

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

DM protein - protein search, using sw model

Run on: October 28, 2003, 15:46:51 ; Search time 86 Seconds

(without alignments)
1002.192 Million cell updates/sec

Title: US-09-856-327-2_COPY_76_618

Perfect score: 2901

Sequence: 1 NAEGETAVPVVPGYHKNEI.....IINTLKGGTDGKNTGHRNL 543

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 19Jun03.*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length DB ID	Description
1	2901	100.0	618 22	Lyophyllum shimeji
2	1142.5	39.4	623 20	Coriolus versicolor
3	1142.5	39.4	623 23	Pyrenopeziza p
4	1136	39.2	633 21	Pleurotus cornucop
5	1128.5	38.9	623 17	Pyrenopeziza
6	1119	38.6	622 22	Trametes hirsuta p
7	1024.5	35.3	566 20	An antitumour prot
8	976.5	33.7	564 21	T. matsutake pyran
9	976.5	33.7	564 21	Trichoderma deri

10	192	6.6	51	20	AAW87541
11	177	6.1	67	21	AAE10459
12	177	6.1	67	21	AAE10459
13	173.5	6.0	539	23	AAU97826
14	170	5.9	1209	22	AAU97826
15	168	5.8	34	22	AAU97826
16	162	5.6	30	22	AAU97826
17	159	5.5	615	20	AAU97826
18	157	5.4	544	20	AAU97826
19	142	4.9	29	22	AAU97826
20	140.5	4.8	562	23	AAU97826
21	130.5	4.5	58	20	AAU97826
22	128.5	4.4	657	21	AAU97826
23	128.5	4.4	675	21	AAU97826
24	128.5	4.4	748	21	AAU97826
25	121	4.2	24	22	AAU97826
26	115	4.0	50	20	AAU97826
27	114.5	3.9	704	20	AAU97826
28	114	3.9	1012	23	AAU97826
29	110	3.8	551	21	AAU97826
30	110	3.8	1498	23	AAU97826
31	108	3.7	35	20	AAU97826
32	106.5	3.7	775	23	AAU97826
33	106.5	3.7	778	20	AAU97826
34	106	3.7	546	20	AAU97826
35	105	3.6	609	22	AAU97826
36	105	3.6	833	22	AAU97826
37	104.5	3.6	1356	22	AAU97826
38	104.5	3.6	1876	18	AAU97826
39	104.5	3.6	1876	22	AAU97826
40	104.5	3.6	1876	22	AAU97826
41	103	3.6	437	22	AAU97826
42	103	3.6	1688	22	AAU97826
43	102.5	3.5	61	20	AAU97826
44	102.5	3.5	352	22	AAU97826
45	102.5	3.5	1686	19	AAU97826

ALIGNMENTS

RESULT 1
AAU97035
ID AAU97035 standard; Protein: 618 AA.
XX
AC AAU97035;

DT 20-JUL-2001 (first entry)

XX Lyophyllum shimeji antibacterial protein.

DE Fungus; antibacterial; antibiotic; plant pathogen; bacterial infection;
XX Pyricularia oryzae; Rhizoctonia solani; rice pathogen.

XX Lyophyllum shimeji.

XX WO200121657-A1

PD 29-MAR-2001

XX 20-SEP-2000; 2000WO-JP06404.

XX 21-SEP-1999; 99JP-0267238.

XX (NISR) JAPAN TOBACCO INC.
XX (NORQ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.

XX Takakura Y, Kuwata S, Inoue Y;

XX WPI; 2001-281598/29.

XX N-PSDB; AAF99980.

XX Antibacterial protein and encoded gene isolated from Lyophyllum

PT shimeji, with activity against plant pathogenic bacteria, applicable in
 PT agriculture e.g. rice cultivation at low concentration, produced at low
 XX cost on large scale
 XX
 XX Claim 3; Page 43-45; 52pp; Japanese.
 CC The present sequence is an antibacterial protein from the fungus
 CC *Lyophyllum shimeji*. The protein was obtained from a fraction prepared
 CC by extracting *Lyophyllum shimeji* with water and subjecting the extract
 CC to ammonium sulphate precipitation. The protein inhibits the growth
 CC of the plant pathogenic bacteria *Pycularia orizae* and *Rhizoctonia*
 CC *solani* at a relatively low concentration. *P. orizae* and *R. solani* are
 CC causative of the two major diseases of rice. The protein contains
 CC components of 70 kDa and 65 kDa, as determined by SDS-PAGE.
 CC The antibacterial protein can be produced at low cost on a large scale.
 XX
 SQ Sequence 618 AA;
 Query Match 100.0%; Score 2901; DB 22; Length 618;
 Best Local Similarity 100.0%; Pred. No. 5.6e-266;
 Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NAEGETAVYVPGYHKKNEIEFQKIDRFVNVIKALQOVSVVRNQNVPTLDPGMSAP 60
 DB 76 NAEGETAVYVPGYHKKNEIEFQKIDRFVNVIKALQOVSVVRNQNVPTLDPGMSAP 135
 QY 61 PGSSAINGKPHOREFENISAEAVTRGVGMSTHTCSTPRIHPHMEISLPGIRPKLSN 120
 DB 136 PGSSAINGKPHOREFENISAEAVTRGVGMSTHTCSTPRIHPHMEISLPGIRPKLSN 195
 QY 121 DPADDKWNELYSAEERLIGTSTKEFDESIRHTLVLSLQDAYKDRQIRFRLPLACHR 180
 DB 196 DPADDKWNELYSAEERLIGTSTKEFDESIRHTLVLSLQDAYKDRQIRFRLPLACHR 255
 QY 181 LKNAPEVWHSANLFSIYNDKQKLFLLTNHRCRTRIALTGGYEKKIGAAEVRNLL 240
 DB 256 LKNAPEVWHSANLFSIYNDKQKLFLLTNHRCRTRIALTGGYEKKIGAAEVRNLL 315
 QY 241 ATRNPSSQDSYIMAKVYVLASGAINPQILYNSGFGSLQVTPRNDSLIPNLGRYITEQ 300
 DB 316 ATRNPSSQDSYIMAKVYVLASGAINPQILYNSGFGSLQVTPRNDSLIPNLGRYITEQ 375
 QY 301 MAFQCVILRQEFVDSVRDDPYGLPWKEVAQAHTAKNPTDALPIPRDPBPQVTTPTTEE 360
 DB 376 MAFQCVILRQEFVDSVRDDPYGLPWKEVAQAHTAKNPTDALPIPRDPBPQVTTPTTEE 435
 QY 361 HPWHTQIHRDAFSGAVGPEVDSRVIVDLRFWFGATDPEANNLLVFQNDVQDYSMPQPTF 420
 DB 436 HPWHTQIHRDAFSGAVGPEVDSRVIVDLRFWFGATDPEANNLLVFQNDVQDYSMPQPTF 495
 QY 421 RYRSTASNVRAKQKMDCEVANSGLGYLPTSPQPMDFGLALHLAGTTRIGFDKATTV 480
 DB 496 RYRSTASNVRAKQKMDCEVANSGLGYLPTSPQPMDFGLALHLAGTTRIGFDKATTV 555
 QY 481 ADNNLSVWDFANLVAGNGTIRTCFGENPTLTSNCHAIKSARSILNTLKGTDCKNTGEH 540
 DB 556 ADNNLSVWDFANLVAGNGTIRTCFGENPTLTSNCHAIKSARSILNTLKGTDCKNTGEH 615
 QY 541 RNL 543
 DB 616 RNL 618
 RESULT 2
 ID AAW94308
 AC AAW94308; standard; protein; 623 AA.
 XX
 XX
 DT 08-APR-1999 (first entry)
 XX
 DE Coriolus versicolor pyranose oxidase.
 XX

KW Coriolus versicolor; modified: pyranose oxidase; heat stability;
 KW pH stability; glucose; glucosone.
 XX Coriolus versicolor.
 XX
 XX JF11009271-A.
 XX
 XX 19-JAN-1999.
 PD
 XX 24-JUN-1997; 97JP-0181865.
 PF
 XX 24-JUN-1997; 97JP-0181865.
 PR
 XX (KIKK) KIKKOMAN CORP.
 PA
 XX WPI; 1999-145891/13.
 DR
 XX
 PT New modified pyranose oxidase gene and protein - useful for
 PT efficient recombinant production of pyranose oxidase with excellent
 PT Km, heat and pH stability
 PT
 XX Claim 1; Page 7-9; 10pp; Japanese.
 PS
 XX The present sequence represents pyranose oxidase derived from Coriolus
 CC versicolor, which can be modified by adding, deleting or replacing at
 CC least one amino acid (preferably where the amino acid at position 542
 CC is replaced) and still retains pyranose oxidase activity. The modified
 CC pyranose oxidase with the amino acid at position 542 replaced, has the
 CC following properties: (a) Action: It oxidises glucose to glucosone;
 CC (b) Substrate specificity: It reacts specifically with glucose and also
 CC reacts with galactose, L-sorbose, D-xylose and 1,5-anhydro-D-glucitol;
 CC (c) Stable pH: 3.5 to 11.0 at 50 degrees Celsius for 30 minutes; (d)
 CC Optimum pH: Near 6.5; (e) Optimum temperature: Near 55 degrees Celsius;
 CC and (f) Temperature stability: Stable up to 55 degrees Celsius. The
 CC present invention also describes a method for the preparation of a
 CC modified pyranose oxidase. The method efficiently prepares a modified
 CC pyranose oxidase with excellent Km value, heat stability and
 CC pH stability.
 XX
 SQ Sequence 623 AA;

Query Match 39.4%; Score 1142.5; DB 20; Length 623;
 Best Local Similarity 45.1%; Pred. No. 5.4e-99;
 Matches 251; Conservative 69; Mismatches 163; Indels 73; Gaps 15;
 QY 13 GYHKKNEIEFQKIDRFVNVIKALQOVSVVRNQNVPTLDPGMSAPPGSSAISNGK 72
 DB 87 GAHKNTVEYQKNDKFNVIQGLMSVSVVNTLVITDITLSPTSWQA--SSFFVRNGSNP 144
 QY 73 HQREFENLSAEAVTRGVGMSTHTCSTPRIHPHMEISLPGIRPKLSNDPAE-DDKEWNE 131
 DB 145 EQPLNLSQAVTRVVGGMSTHTCATPRFREQ-----RPLLKDDQDADAEDR 197
 QY 132 LYSEAEERLIGTSTKEFDESIRHTLVLSLQDAYKDRQIRFRLPLACHRLKNAPEYVEWH 191
 DB 198 LYTKASVYFKTGTQDKESIRHNLVNLKLAEEYKG-ORDFQDIPLATR--RSETFFVWS 254
 QY 192 SAENLF--HSIYNDKQKLFLLTNHRCRTRIALTGGYEKKIGAAEVRNLLATRNPSQL 249
 DB 255 SANTVFDLQNRPNTPDAPNERFNLFPVACERV-----VRN-----TSNSEI 295
 QY 250 DSY-----IMAKVYVLASGAINPQILYNSGFGSLQVTPRNDSLIPNLGRY 295
 DB 296 ESLHIDLISGRFEIKADVFLTAGAVHNAQLVNSGFGQLGRPDNPANPQLPLSLG 355
 QY 296 ITEQPNFAQCVILRQEFVDSVRDDPY--GLP-----MWEKVAQH 333
 DB 356 ITEQSLVFCQTVNSTELIDSVKSDMIRGNPDLGYSVTVTPGAETNKHGDMNEKVRKH 415
 QY 334 IAKNPTDALPIPRDPBPQVTTPTTBEHPWHTQIHRDAFSGAVGPEVDSRVIVDLRFWFG 393
 DB 416 MMQHQEDPLPIPRDPBPQVTTTLFQSPHPWHTQIHRDAFSGAVGPEVDSRVIVDLRFWFG 475

QY 394 ATDPANLLVQNDVQGVSMPOPTFRYR--PSTASNVARKMMADMCCEVASNLGGYLPT 452
 Db 476 RTEPKEENKLFWSKIDTDTYNMPOPTDFRPPAGRTSKEAEDMMTDMCVMSAKIGGLPG 535
 QY 453 SPQFMDPGLALHLAGTTRIGFDKA--TTVADNNSLVDFANLYVAGNGTIRGTGGENPT 510
 Db 536 SLQPFMEFGLVHLGGTHRGFDQEDKCCVNTDSRVFGFNFLGGCGNIPTAYGANPT 595
 QY 511 LTSMCHAIKSARSIIIN 526
 Db 596 LTAMSLAIKSCYIKN 611

RESULT 3

AAO15842
 ID AAO15842 standard; Protein; 623 AA.

XX AC AAO15842;

DT 19-DEC-2002 (first entry)

XX DE

XX PYranose oxidase protein.

XX Recombinant polypeptide display method; signal peptide; pyranose oxidase;

XX transmembrane linker; transporter domain; autotransporter; enzyme;

XX protease recognition site; catalytic biomolecule design;

XX cell factory design; adrenotoxin; mitochondrial target sequence.

XX Unidentified.

XX WO200270645-A2.

XX 12-SEP-2002.

XX 01-MAR-2002; 2002WO-EP02246.

XX 02-MAR-2001; 2001EP-0105129.

XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX Jose J, Hannemann F, Bernhardt R;

XX WPI; 2002-732791/79.

XX Displaying recombinant functional polypeptide containing prosthetic

XX group or several subunits, on host cell surface using the transporter

XX domain of an autotransporter which is fused to a signal peptide -

XX Disclosure; Fig 14; 80pp; English.

XX The invention comprises a method of displaying a recombinant polypeptide

XX containing a prosthetic group on the surface of a host cell. The method

XX involves transforming a host cell with a nucleic acid construct that

XX encodes a signal peptide, the polypeptide to be displayed, a

XX transmembrane linker, a transporter domain to be displayed, a

XX optionally a protease recognition site. The method of the invention is

XX useful for displaying a recombinant polypeptide containing a prosthetic

XX group on the surface of a host cell, where the prosthetic group comprises

XX an inorganic component such as a metal (e.g. cobalt, nickel, manganese,

XX copper or iron) containing group. The method of the invention has

XX applications in evolutive design of catalytic biomolecules or for new

XX whole cell factories. The present amino acid sequence represents a

XX pyranose oxidase protein.

XX Sequence 623 AA;

XX Query Match 39.4%; Score 1142.5; DB 23; Length 623;

XX Best Local Similarity 45.1%; Pred. No. 5,4e-99;

XX Matches 251; Conservative 69; Mismatches 163; Indels 73; Gaps 15;

XX 13 GYHKNETFQKIDRFNVNITKGALQQVSVPRNVNPTLDPGAWSPGSSAISNGKNP 72

XX 87 GAHKNTVEYQKIDKFNVIQGLMSVSVFNTLVIDTLSTSQQA--SSFFVRNGSNP 144

QY 73 HOREFENLSAEAVTRGVGGMSTHTCTSTPRIHPMBESLPGLGRPKLSNDPAE--DDKEWNE 131
 Db 145 EQPLRLNSQAVTRVVGGMSTHTCATPRDREQ-----RPLLVKDDQDADDAEWR 197
 QY 132 LYSAEERLGTSTKBFDESIRHTLVLSLQDAYKDRQIRFRPLPLACHRLKNAPEYVEWH 191
 Db 198 LYTAKSSYFKTGTDQFKESIRHNLVLNKLAEYKG--QRDFQIPLAATR--RSPTTFVWS 254
 QY 192 SAENLF--HSIYNDDKQKLFLLTNHRCRTRALTGGYEKKIGAAEVRNLLATRNPSQL 249
 Db 255 SANTVFELQNRPNTPDAPNERFNLFPVACERV-----VRN-----TSNSEI 295
 QY 250 DSY-----IMAKYVVLASCAIGNPQILYNSGFSGL--QVTPRN--DSLIPNLGRY 295
 Db 296 ESLHIHDLISGRFEIKADVFLTAGAVHNAQLVNSGFGQLGRPDPAFPQLPLSLGSY 355
 QY 296 ITEQPMACQIVLRQEFVDSVRDDPY--GLP-----MWKEAVAQ 333
 Db 356 ITEQSLVFCQVMSTELIDSVKSDMIIRGNPGDLGYVTVTPCAETNKHDPWNEKVKH 415
 QY 334 IAKNPTDALPIPRDEPOVTPPTEEHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRWF 393
 Db 416 MMQHOEDPLPIPFEDPEPOVTTLFQPSHPWHTQIHRDAFSYGAVQOSIDSLIVDMRFFG 475
 QY 394 ATDPEANLLVFQNDVQGVSMPOPTFRYR--PSTASNVARKMMADMCCEVASNLGGYLPT 452
 Db 476 RTEPKEENKLFWSKIDTDTYNMPOPTDFRPPAGRTSKEAEDMMTDMCVMSAKIGGLPG 535
 QY 453 SPQFMDPGLALHLAGTTRIGFDKA--TTVADNNSLVDFANLYVAGNGTIRGTGGENPT 510
 Db 536 SLQPFMEFGLVHLGGTHRGFDQEDKCCVNTDSRVFGFNFLGGCGNIPTAYGANPT 595
 QY 511 LTSMCHAIKSARSIIIN 526
 Db 596 LTAMSLAIKSCYIKN 611

RESULT 4

AAAY52700

ID AAY52700 standard; Protein; 633 AA.

XX AC AAY52700;

XX DT 07-MAR-2000 (first entry)

XX DE Pleurotus cornucopiae antitumour protein.

XX KW Antitumour; cancer; tumour; treatment; expression; tumour suppressor;

XX PW53; pBR; ss.

XX OS Pleurotus cornucopiae.

XX PN JP11315096-A.

XX PD 16-NOV-1999.

XX PF 07-AUG-1998; 98JP-0216349.

XX PR 08-AUG-1997; 97JP-0215311.

XX PR 02-MAR-1998; 98JP-0066176.

XX PA (NEWF-) NEW FOOD CREATION GIJUTSU KENKYU KOMIAI.

XX WPI; 2000-058170/05.

XX DR N-PSDB; AA46411.

XX An antitumour protein derived from Pleurotus cornucopiae and its gene

XX useful for treatment of cancer including those caused by abnormal

XX expression of cancer inhibitory gene (e.g. p53 and pBR)

XX Claim 3; Page 13-15; 23pp; Japanese.

CC The invention relates to a novel antitumour protein extracted from
 CC fruiting bodies of the fungus *Pleurotus cornuopae*. The protein and
 CC nucleotides encoding it are useful for the treatment of cancer,
 CC including those caused by abnormal expression of tumour suppressor
 CC genes such as p53 and pBR. This sequence represents the antitumour
 CC protein.
 XX SQ Sequence 633 AA;
 Query Match 39.2%; Score 1136; DB 21; Length 633;
 Best Local Similarity 45.3%; Pred. No. 2.3e-98;
 Matches 248; Conservative 70; Mismatches 175; Indels 54; Gaps 15;
 QY 13 GYHKNEIEFQKIDRFNVNIGALQOVSVPRNQVNPVLPDPCAWAPPGSSAISNGKNP 72
 DB 96 GSHKQNTVEYQKNDKFNVIQGLMPVSVPRNKYVADTLSPASWQA--STIFVRNGANP 153
 QY 73 HOREPENLSAEAVTRGVGGMSTHTCTSTPRIHPPMESLPGIGRPKL-SNDPAEDDKEMNE 131
 DB 154 EQDFTNLSCQAVTRVVGGMATHWTATPRFKS-----EPKLVKDDSDADALEWER 206
 QY 132 LYSEAEERLIGTSKEFDESIRHTLVLSLQDAYKDRQIRFRPLPLACHRLKNAPEYVEWH 191
 DB 207 LYDIAESFVKTGHQFDQSIHRNLVLEKLQESYAG-QRGFEQIPLAAQRTN--PRFVENS 263
 QY 192 SAENLF--HSIYNDKQKCLFTLLTNHRCRLALTGGEYKKGAAAEVRNLLATRNPSQL 249
 DB 264 SAHTVFDLENRPNADDEKGRNLFPAVVCERVTRDSDRKIDIEVHDLISG----- 315
 QY 250 DSI-IMAKVYVLASGAIGNPOILYNSGFSGLQVTPRND---SLIPNLGRYTEQPMFAC 304
 DB 316 DRYKVKADVFILCSGAVHNPOILVNSGF-GRMGQDSSLPPTLLPYLSYITEQTLTFC 374
 QY 305 QIVLRQEFVDSVRDDP--YGLP-----NMKEAQAQHIKAKNPTDAL 342
 DB 375 QTVFSTELNVLKSDMIIVGTPGPDYVTFPTDPSNKHPPNWNNEKVKHMQHEDPL 434
 QY 343 PIFRPEPEQVTPFTTEHPHWTQIHRDAFSYGAVGSDSRVIVDLRMFGATDPEANLL 402
 DB 435 PFLDDPEQVTTLFQDTHPWHQIHRDAFSYGAVAESIDSLVLDVDMRFFGRTEPVEENK 494
 QY 403 LVPQNDVQDGYSMPOPTFRYR--PSTASNVRAKRMADMCMCEVSNLGYLPTSPQPMWDPG 461
 DB 495 LWFESKQITDAYNPQPTFSRPFQGRTAQAEELMADMCMSTKVGGLPGSYVQPMFAG 554
 QY 462 LALHLACTTRIGFDKATTVA--DNNSLVMDFANLYVAGNGTIRTFGENPTLTSMCHAIK 519
 DB 555 LVLHLGCTHRMGFDEADKACVDNTSKVFCMENLFLGGCGNICTAYASNPTLTAVALAIR 614
 QY 520 SARSIIN 526
 DB 615 SKYIRN 621

RESULT 5

AAR99628
 ID AAR99628 standard; Protein; 623 AA.

XX AC AAR99628;

XX DT 27-NOV-1996 (first entry)

XX XX Pyranose oxidase.

XX DE Pyranose oxidase; glucose; oxidation; glucanase; assay; diabetes;

XX KW marker; diagnosis; 1,5-anhydro-D-sorbitol.

XX OS Coriolus versicolor.

XX PN DE19545780-A1.

XX PD 13-JUN-1996.

PF 07-DEC-1995; 95DE-1045780.
 XX 24-MAY-1995; 95JP-0124835.
 PR 07-DEC-1994; 94JP-0304086.
 XX (KIKK) KIKKOWAN CORP.
 XX Kawai G, Koyama Y, Minamihara T, Nishimura I, Okada K;
 PI Suzuki M;
 XX WPI; 1996-278990/29.
 DR N-PSDB; AAT34420.
 XX DNA encoding protein with pyranose oxidase activity at neutral pH -
 PT useful for the determination of glucose in body fluids or foods, or
 PT 1,5-anhydro-D-sorbitol used as marker for diabetes diagnosis
 XX Claim 1; Page 14-16; 22pp; German.
 XX The present sequence encodes a protein isolated from *Coriolus versicolor*,
 CC which has the enzyme activity of pyranose oxidase (PO). The PO oxidises
 CC glucose to gluconone and has an optimum pH of 7-7.5. It has a mol. wt. of
 CC 29000 (determined by gel filtration) and is stable at around 50deg.C.
 CC PO can be used for measurement of glucose in, e.g. foods or body fluids,
 CC or 1,5-anhydro-D-sorbitol which is an important marker used in the
 CC diagnosis of diabetes.
 XX SQ Sequence 623 AA;
 Query Match 38.9%; Score 1128.5; DB 17; Length 623;
 Best Local Similarity 45.0%; Pred. No. 1.1e-97;
 Matches 250; Conservative 68; Mismatches 165; Indels 73; Gaps 15;
 QY 13 GYHKNEIEFQKIDRFNVNIGALQOVSVPRNQVNPVLPDPCAWAPPGSSAISNGKNP 72
 DB 87 GAKKQNTVEYQKNDKFNVIQGLMSVSPVNTLVITLSPTSWQA--SSFFVRNGSNP 144
 QY 73 HOREPENLSAEAVTRGVGGMSTHTCTSTPRIHPPMESLPGIGRPKLSNDPAE--DDKEWNE 131
 DB 145 EQDFTNLSCQAVTRVVGGMSTHTCTATPRFREQ-----RPLLVKDDQDADDAEWDR 197
 QY 132 LYSEAEERLIGTSKEFDESIRHTLVLSLQDAYKDRQIRFRPLPLACHRLKNAPEYVEWH 191
 DB 198 LYTAKESYFKTQDKPKESIRHNLVNLKLAEEYKG-QRDFQIPLAATR--RSPTFVEMS 254
 QY 192 SAENLF--HSIYNDKQKCLFTLLTNHRCRLALTGGEYKKGAAAEVRNLLATRNPSQL 249
 DB 255 SANTVFDLQNRPNTPDAPNERFNLFPVAVACERV-----VRN-----TSNSEI 295
 QY 250 DSI-----IMAKVYVLASGAIGNPOILYNSGFSGL--QVTPRN--DSLIPNLGRY 295
 DB 296 ESLHIHDLISGDRFEIKADVFLITAGAVHNAQLLVNSGFGQLGRPDNPANPQLPLSURY 355
 QY 296 ITEQPMFACQIVLRQEFVDSVRDDPY--GLP-----NMKEAQAQHIKAKNPTDAL 333
 DB 356 ITEQSLVFCQTVNMTSLIDSVKSDMIIRGNPGDLGYSVTVTPCAETNKHFDWNEKVKHN 415
 QY 334 IAKNPTDALPIFRDPEQVTPFTTEHPHWTQIHRDAFSYGAVGSDSRVIVDLRMFG 393
 DB 416 MMHQHEDPLPIPEDPEQVTTLPQSPHPHWTQIHRDAFSYGAVQVQITDRLIVDMRFFG 475
 QY 394 ATDPEANLLVQNDVQDGYSMPOPTFRYR--PSTASNVRAKRMADMCMCEVSNLGYLPT 452
 DB 476 RTEPKEENKLFWSDKITDYNMPOPTDFRFPAGRTSKAEADMMTDMCVNWKIGGLPG 535
 QY 453 SPQFMDPGLALHLACTTRIGFDKA--TTVADNNSLVMDFANLYVAGNGTIRTFGENPT 510
 DB 536 SLQFQMEPGLVHLGCTHRMGFDEADKCCVNTDSRVFGKFLFLGGCGNICTAYGANPT 595
 QY 511 LTSMCHAIKSARSIIN 526
 DB 596 LTAMSLAIKSCYEIKN 611

```

RESULT 6
ID AAB48832
DB AAB48832 standard; Protein; 622 AA.
CX AAB48832;
CX AAB48832;
CX 13-MAR-2001 (first entry)
CX Trametes hirsuta pyranose oxidase.
CX Pyranose oxidase; expression construct; recombinant production;
CX monosaccharide oxidation; 2-keto derivative;
CX hydrogen peroxide production.
CX Trametes hirsuta.
CX US6146865-A.
CX 14-NOV-2000.
CX 05-MAY-1999; 99US-0305381.
CX 08-JUN-1998; 98DK-0000774.
CX 10-JUN-1998; 98US-0088724.
CX (NOVO) NOVO NORDISK AS.
CX Schneider P, Christensen S, Lassen SF;
CX WPI; 2001-049055/06.
CX N-PSDB; AAC87518, AAC87519.
CX Novel nucleic acid molecule encoding polypeptide having pyranose
CX oxidase activity used to design oligonucleotide probes to identify and
CX clone DNA encoding the polypeptide from different genera or species -
CX Claim 5; Fig 1; 20pp; English.
CX The invention relates to nucleic acids (e.g., AAC87518, AAC87519) which
CX encode Trametes hirsuta pyranose oxidase (AAB48832). The invention also
CX relates to expression constructs, expression vectors and recombinant
CX cells comprising pyranose oxidase nucleic acid sequences, and the
CX recombinant production of Trametes hirsuta pyranose oxidase. Pyranose
CX oxidase catalyses the oxidation of several monosaccharides in the
CX pyranose form at position C2 to produce 2-keto derivatives with the
CX release of hydrogen peroxide. Nucleic acids encoding Trametes hirsuta
CX pyranose oxidase may be used to produce the enzyme and to design
CX oligonucleotide probes to identify and clone genomic pyranose oxidase
CX cDNA or genomic DNA from different genera or species of microorganisms
CX (fungi or bacteria). The present sequence represents pyranose oxidase
CX from the fungus Trametes hirsuta.
CX Q Sequence 622 AA;
Query Match 38.6%; Score 1119; DB 22; Length 622;
Best Local Similarity 44.6%; Pred. No. 9.1e-97;
Matches 251; Conservative 69; Mismatches 183; Indels 60; Gaps 15;
13 GYKKNEIEFQKIDRFVNVIKGALQOVSVVRNQNVPTLDPGAWSPGSSAITSNGKNP 72
87 GSHKENVVQKNDKFNVIQGLMPVSVVNTWVDTLSPASQA--STFFVRNGANP 144
73 HQREFENLSAEAVTRGVGGSTHTWCTSPRIHPMPESLPGIGRKL-SNDPAEDDKWNE 131
145 EQDFLRNLGQAVTRVVGGMSTHTWCATPFK-----LQPELLVKNDSKADDAEWDR 197
132 LYSEAEELGTSIKPEPESRHTVLVLSLQDAYKDORIRFRLPLACHRLKNAPYVEWH 191
198 LYKAEYSFKTGTTFQAEISRHNVLVKKLOEYKGV-VRDFOQIPLAATR--QSPTFFVWS 254
192 SAENLF--HSIYNDKOKKLTLLTNHRCRLALTGGEYKKGAEVRNLLATNPSSQL 249
255 SAHTVFDLENRENKDKAPKQRFNLFPFAVACTNVRNRNANSEIVG-LDVRDLHGKGSIT--- 310
250 DSYIMAKUYVILASGAIGNPQILYNSGFSGLQVTPRND-----SLIPNLGRVITSPQMAP 303
311 ---IKAKVYIILTAGAVHNAQLAASGFGQL---GRDPAPAKPLPSLLPYLGHITQTLVF 364
304 COIVLURQEFVDSVRDDP--YGLP-----WKEAVAQHIKKNPTDA 341
365 CQTVNSTELINSVTADMTIVGKPGHPDYSVTVTPGNPNKNKIPDMWNEKVKCHMDHQEDP 424
342 LPFPRDPEPQVTPPTTEHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRFWFGATDPEANN 401
425 LPFPEDEPQVTTLFQATHPWHTQIHRDAFSYGAVQOSIDSRLIVDWRFFGRTPEKSEN 484
402 LLVFQNDVQDGYSMPOPTFRYRPSTASNVARKMMADNCEVASNLGGVLPSPPOFMDPG 461
485 KLFESDKITDAYNLROPTDFR--FPGGREAEADMTDMCVMSAKITGGFPGSYPOFMEPG 542
462 LALHLAGTTRIGFDKAT--TVADNNSLVMDFANLYVAGNGTIRTCFGENPTLTSCHAIK 519
543 LVHLGGTHRMGFDEKADKCCVDTDSRVFGFNULFLGCGGNIPTAYAAANPTLTAMSLAIK 602
520 SARSIINTLKGTDGKNTGEHRN 542
603 SCEVI---KKNFEPSPNPVKHN 622
RESULT 7
AAW87531
ID AAW87531 standard; Protein; 566 AA.
CX AAW87531;
CX AAW87531;
CX 26-FEB-1999 (first entry)
CX An antitumour protein.
CX Antitumour protein; Tricholoma matsutake.
CX Tricholoma matsutake.
CX JPI0313876-A.
CX 02-DEC-1998.
CX 13-FEB-1998; 98JP-0031452.
CX 13-FEB-1997; 97JP-0029275.
CX (MOMO-) MOMOYA KK.
CX (NORO) NORINSUISANSHO SHOKUHN SOGO.
CX WPI; 1999-074153/07.
CX N-PSDB; AAW83626.
CX An anti-tumour protein - prepared by culture of host cell
CX transformed by vector containing base coding sequence
CX Claim 1; Page 7-8; 15pp Japanese.
CX The present sequence encodes an antitumour protein, and is isolated
CX from Tricholoma matsutake.
CX Sequence 566 AA;
Query Match 35.3%; Score 1024.5; DB 20; Length 566;
Best Local Similarity 43.0%; Pred. No. 7.3e-88;
Matches 232; Conservative 77; Mismatches 180; Indels 51; Gaps 12;
2 AEGGTAVPYVPGVHKNEIEFQKIDRFVNVIKGALQOVSVVRNQNVPTLDPGAKSAPP 61
62 AEIGSQDNFVIGAHKNSIKFQKIDRFVNVIKGALQOVSVVRNQNVPTLDPGAKSAPP 120

```

Qy	62	GSSA----	ISGNKNPHQREFFENLSAEAVTRGVGGGNSTHTWCSTPRIHPHPPMESLPGGIRPK	117
Db	121	IDPAEQQLVIMGHNPNQEAGLNLP	GSASVTRTVGGMATHWTCACPTPHD-----EE	170
Qy	118	LSNDPAEDDKENNELSYSEAERLIGSTKEPDESRHTVLVLSLODAYKQRI	FRPLPLA	177
Db	171	RVNVPV-DKQEFDA	LLERAKTLNVLVHSDQYDDSRQIVKETLQOOTL-DASRGVTTPLG	228
Qy	178	CHRLKNAPSYVEWHSAENLFHSIYNDDKQKFLTLLTNHRC	TRLALTGGYEKKIGAAEVR	237
Db	229	VERRTDNPIYVTWTCAD---	TVLGDVPKSPRFALVTETRVTKLIVSETNPTQVVAALLR	284
Qy	238	NLLATRNPSQLDSYIMAKVYVLASGAIGNPQILYNSGFSGLQVTPRND	SLIPNLGRYIT	297
Db	285	NL-----NTSN--DELVWAKSFVACGAVCTPQILWNS----	NIRPY-----ALGRYLS	327
Qy	298	EOPMAFCQILVLRQEFVDSVRDDPYGLPWKAEVAQHIAKNPTD	ALPIPERDEPQVTTFP	357
Db	328	ESQMTFCQILVLRGIVDAIADPR----	FAAKVBAHKKKHGPDVLPPIPHPEPEPQWMPY	383
Qy	358	TEEHPWHTQIHRDAFSYGAVGPEVDSRVIYDLRWF	GATDPEANNLLVF-----QND	408
Db	384	TSDFPMWVQVHRDAFSYGDVGPKADPRVYVDLRF	FGKSDIVEENRVTFGPNPKLREWEAG	443
Qy	409	VODGYSMPQPTFRYRTPSTASVNRARKMADMC	EVASNLGGLYPTSPQPMDFGLALHLHLAG	468
Db	444	VTDITGYMFQPTPHVKRTNADGDRDQRMMNDMTN	VANMLGGLYPGSYPQFMAPGLVLHITG	503
Qy	469	TTRIGCFKATTVADNNSLVMPDFANLYVAGNGTIR	TGFGENPTLTSMCHAIKSARSIINTL	528
Db	504	TTRITGDDQTSVADPTS	KVHNFNNLWVGNGCIPDATACNPTRTSVAYALKGAEAVVNYL	563
		RESULT 8		
		AAB10457		
ID		AAB10457	standard; Protein; 564 AA.	
XX		XX		
AC		AC	AAB10457;	
XX		XX		
DT		DT	11-DEC-2000 (first entry)	
XX		XX		
DE		DE	T. matsutake pyranose oxidase protein.	
XX		XX		
KW		KW	Pyranose oxidase; matsutake mushroom; antibacterial; sugar metabolism;	
XX		XX	diabetes mellitus.	
XX		XX	Tricholoma matsutake.	
OS		OS	JP2000175698-A.	
XX		XX		
PN		PN	27-JUN-2000.	
XX		XX		
PF		PF	16-DEC-1998; 98JP-0357423.	
XX		XX		
PR		PR	16-DEC-1998; 98JP-0357423.	
XX		XX		
XX		XX	(NIBS) JAPAN TOBACCO INC.	
XX		XX		
DR		DR	WPI; 2000-478485/42.	
DR		DR	N-PSDB: AAA71487.	

```
CC activity by being heated at 60 degrees C for 10 minutes in a neutral
CC aqueous solution and in which the above antibacterial activity is
CC inactivated by being heated at 80 degrees C for 10 minutes in a neutral
CC aqueous solution. The invention also describes a method for the
CC determination of pyranose in a sample in which the above reagent for the
CC determination of pyranose is reacted with pyranose in the sample and the
CC hydrogen peroxide formed is reacted with an enzyme to develop a color, a
CC method for diagnosing a disease accompanied by abnormality in sugar
CC metabolism in which the above reagent for the determination of pyranose
CC is reacted with pyranose in the sample and the hydrogen peroxide formed
CC is reacted with an enzyme to develop a color, and a kit for pyranose
CC analysis or the diagnosis of a disease accompanied by abnormality in
CC sugar metabolism. The pyranose oxidase can be used as a diagnostic agent
CC for diabetes mellitus. This sequence represents the Tricholoma matsutake
CC pyranose oxidase protein which is described in the method of the
XX invention.
```

SQ Sequence 564 AA;

Query Match 33.7%; Score 976.5; DB 21; Length 564;
Best Local Similarity 41.3%; Pred. No. 2.6e-83;
Matches 223; Conservative 82; Mismatches 182; Indels 53; Gaps 13

QY 2 AEGTAVPYVGYGHKQNEIEFQKDIDRFVNVIKGALQQVSVPVRNQNYPTLDPGWSAPP 61
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 AEIGSQDNPIVGAHRRNSIKFOKDTDFVNIINGALQPISPSPTYQTLLAFAAW-APP 120
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 62 GSSA-----INSKNPHOREPENLSAEAVTRGVGGMSTHTWCSTPRIHPPMESLPDGRPK 117
Db 121 IDPAEGLVLNGHNPNQEAGLNLPGSATRTVGGMATHWTCACTPHD-----EE 170
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 118 LSNDDPADDEKEWNELYSSEERLICTSKPEDESIRHTLVLRSLQDAYKDORIFRPLPLA 177
Db 171 RVNPNP-VDKQEFDALESRAKTLLNVHSDDYDDSIROI VVKETLOOTL-DASRGVTTLFLG 228
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 178 CHRUKNAPEYYVEHSAENLFHSIYNDQKKLFTLLTNHRCTRALTGGYEKKIGAABEV 237
Db 229 VERRTDNPVIYVTMGAD----TVLGDPVKS PRFVLVTETRVTKIVSETNPVVVAALLR 284
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 238 NLLATRPPSQLDSYIMAKYVYLASGAIGNPOILYNBSFGSLQVTPRNDSLIPNLGRYT 297
Db 285 NL-----NTSN--DELVWAQS FVIAACGAVCTPQLWN S-----NIRPH-----ALGRYLS 327
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 298 EQMAFCQIVLROBFVDSDRDPXGLPWMEAAVQAHIKQNPTDALPIPRDPEPOVTTTF 357
Db 328 EOSTWFQCVLKRSIVDSIATDR ----FAAKEAHKKRPDDVLPITFHFEPOVMILPY 383
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 358 TEEHPWHQTHRDAFSYGAVGEVDSRVIVDLRWFGATDPEANLLVF-----QND 408
Db 384 TSDFPWHVQH R --YAFGDVGPRADPRVVVDLRFPGKSDIVEENRVTFGNPKLRDWEG 441
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 409 VDQYSNMQPQSTRYPYSTASNVAROKMDACEVASNLGGYLPSPFPQM DPGLALHLUAG 468
Db 442 VTDTYGHPQPTFFHKRTNADGDDRDMNM DMTN VANILGGYLPGS YPQFMAPGLAQHITG 501
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 469 TTTRIGFDKATTADDNNSLVMDFANLYVAGGTGIRTGFGENPTLTSMCHAIKARSINTL 528
Db 502 TTRI GTDDQTSVADPTSKVNFNFNLWGNGCIPDATA CNPTRITSVAYALKGA EVSVYL 561
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
AAy81952
ID AAY81952 standard; Protein; 564 AA.
XX AC AAY81952;
XX DT 07-JUL-2000 (first entry)
XX DE Trichoderma derived antifungal protein sequence.
XX KW Antifungal protein; Pyricularia oryzae; Rhizoctonia solani; mushroom;
KW growth inhibitor; plant pathogenic fungi; antibacterial agent;
KW N-terminal fragment.

XX OS Trichoderma matsutake.
 XX XX WO200014242-A1.
 XX PN 16-MAR-2000.
 XX PD 19-AUG-1999; 99WO-JP04441.
 XX XX 08-SEP-1998; 98JP-0270606.
 XX PR (NISR) JAPAN TOBACCO INC.
 XX PA (NORQ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.
 XX PI Takakura Y, Kuwata S, Ohta S;
 XX DR WPI; 2000-256990/22.
 XX DR N-PSDB; AAA07403.
 XX PT Mushroom-derived antibacterial protein against plant pathogenic fungi
 XX PT of rice, with activity and thermal stability, obtainable cheaply on
 XX PT large scale, useful in agriculture
 XX XX Claim 3; Page 41-45; 52pp; Japanese.
 XX PS This sequence is the Trichoderma matsutake antibacterial protein of
 XX CC the invention. The protein has activity against at least Pyricularia
 XX CC oryzae and Rhizoctonia solani, and is obtained from a fraction of an
 XX CC aqueous extract of a mushroom precipitated by the ammonium sulphate
 XX CC precipitation method. The protein has a molecular weight of about 210 kD
 XX CC as determined by the gel filtration method, includes components of about
 XX CC 15 kD and 50 kD in SDS-PAGE, and is stable to heating in an aqueous
 XX CC neutral solution at 60 degrees C for 10 minutes but with loss of
 XX CC antibacterial activity after heating in the solution at 80 degrees C for
 XX CC 10 minutes. The protein is used for inhibiting the growth of plant
 XX CC pathogenic fungi e.g. Pyricularia oryzae and Rhizoctonia solani.
 XX CC It is useful in treating rice plants, and is applicable in agriculture as
 XX CC an antibacterial agent. The protein has activity at relatively low
 XX CC concentrations, and can be produced at low cost on large scale.
 XX X Sequence 564 AA;
 X Q Query Match 33.7%; Score 976.5; DB 21; Length 564;
 X Q Best Local Similarity 41.3%; Pred. No. 2.6e-83;
 X Q Matches 223; Conservative 82; Mismatches 182; Indels 53; Gaps 13;
 Y 2 ABEGTAVPVPGYHKNEIEFQKIDRFVNVIKGALQOVSVPRNQNVPTLPDGAWSAPP 61
 b 62 ABIGSQDNPVIGAHNRNSKFKDXTDKFVNIINGALOPISISPSDYQPTLVAWAW-APP 120
 Y 62 GSSA-----ISNGKNPHQREFENLSAEAVTRGVGMSTHWTCTSPRIHPMPSLPGIGRPK 117
 b 121 IDPABGQLVIMGNHPNQEAGNLPGSAVTRTVGGMATHWTCACTPHD-----EE 170
 Y 118 LSNDPAEDDKENWELYSAEERLIGTKEPDSIRHTLVLSQDAYKDRQRIAPLPLA 177
 b 171 RVNNEPV-DKQEFDALEAKTLNLSHQYDSDSIRQIVVVKETLQOQL-DASRGVTTPLG 228
 Y 178 CHRLKNAPEYVYHSAENLPHSIYNDKQKLTLLTNHCTRLALTGGYEKKIGAEVR 237
 b 229 VERRTDNIYVWTGAD----TVLGDVPKSPRFVLVTETRTVKFIVSETNPTQVVAALLR 284
 Y 238 NLLATRNPSQLDSYIMAKVYVLSAIGNPOILYNSGFSGLQVTPRNDLSLTPNLGRVIT 297
 b 285 NL----NTSN--DELVAQSFVIACGACTPQLWNS-----NIRPH-----ALGRVLS 327
 Y 298 EQPMAFQCIULROEVRDSDYDPPGLPWWKEVAOHIAKNPTDALPIFRDPEQVTPPF 357
 b 328 EQSMTFCQVLKRSIVDSIATDPR----FAAKVEAHKKKHDPDVLPIPFHEPEQVMIPY 383
 Y 358 TEHPHWTQIHRDAFSGYAGVGPEDSVRVLDLRFEGATDPEANNLLVF-----QND 408
 b 384 TSDFPFWQVHR--YAFGDVGKADPRVVVDLRFEGKSDIVEENRVTFGPNKLRDWEAG 441

409 VQDGYSNPQPTFRYRSTASNVARQKMDMCEVSNLGGYLPSTPQFMDPGLALHLAG 468
 442 VTDYGMQPTFHVKRTNADGDRDQRMNDMTNVANILGGLPGSYLPQFWAPGLAQHITG 501
 469 TTRIGFDKATTVADNNSLVWDFANLVVAGNGTTRTGTGNGENPTLTSMCHAIKSARSINTL 528
 502 TTRIGTDDQTSVADPTSKVHNFNLMVGGNGCIPDATACTNPTRTSVAYALKGAEEVVSYL 561

RESULT 10
 AAW87541
 ID AAW87541 standard; Peptide; 51 AA.
 XX AC AAW87541;
 XX XX 26-FEB-1999 (first entry)
 XX DE Peptide derived from an antitumour protein.
 XX KW Antitumour protein; Tricholoma matsutake.
 XX OS Tricholoma matsutake.
 XX PN JP10313876-A.
 XX PD 02-DEC-1998.
 XX PF 13-FEB-1998; 98JP-0031452.
 XX PR 13-FEB-1997; 97JP-0029275.
 XX XX (MOMO-) MOMOYA KK.
 XX PA (NORQ) NORINSUISANSHO SHOKUHN SOGO.
 XX DR WPI; 1999-074153/07.
 XX PT An anti-tumour protein - prepared by culture of host cell
 XX PT transformed by vector containing base coding sequence
 XX PS Claim 5; Page 12; 15pp; Japanese.
 XX CC The present sequence represents a peptide derived from an antitumour
 XX CC protein. Nucleic acid encoding the antitumour protein is isolated
 XX CC from Tricholoma matsutake.
 XX SQ Sequence 51 AA;
 X Q Query Match 6.6%; Score 192; DB 20; Length 51;
 X Q Best Local Similarity 62.7%; Pred. No. 1.5e-10;
 X Q Matches 32; Conservative 6; Mismatches 13; Indels 0; Gaps 0;
 Y 330 VAQHTAKNPTDALPIFRDPEQVTPTEHPHWTQIHRDAFSGYAGVPE 380
 b 1 VEAKKKHFDVLLPIPFHEPEQVMIPYTSDFPWHVQVHRDAFSGYDVGPK 51

RESULT 11
 AAB10459
 ID AAB10459 standard; Protein; 67 AA.
 XX AC AAB10459;
 XX XX 11-DEC-2000 (first entry)
 XX DE T. matsutake pyranose oxidase protein N-terminal fragment #2.
 XX KW Pyranose oxidase; matsutake mushroom; antibacterial; sugar metabolism;
 XX KW diabetes mellitus.
 XX OS Tricholoma matsutake.
 XX PN JP2000175698-A.

XX 27-JUN-2000.
 PD 16-DEC-1998; 98JP-0357423.
 XX 16-DEC-1998; 98JP-0357423.
 PF 16-DEC-1998; 98JP-0357423.
 PR (NISR) JAPAN TOBACCO INC.
 XX WPI; 2000-478485/42.
 DR A reagent containing pyranose oxidase for the determination of pyranose
 XX Claim 3; Page 19; 23pp; Japanese.
 PT This invention describes a novel reagent containing pyranose oxidase for
 XX the determination of pyranose which can be prepared from a fraction
 CC precipitated from an aqueous extract of matsutake mushroom by ammonium
 CC sulfate precipitation and has an antibacterial activity against at least
 CC Pyricularia oryzae or Thanatephorus cucumeris and has a M.W. of ca.
 CC 210 kD by gel filtration and shows the presence of components of ca.
 CC 50 kD and ca. 15 kD by SDS-PAGE and maintains the above antibacterial
 CC activity by being heated at 60 degrees C for 10 minutes in a neutral
 CC aqueous solution and in which the above antibacterial activity is
 CC inactivated by being heated at 80 degrees C for 10 minutes in a neutral
 CC aqueous solution. The invention also describes a method for the
 CC determination of pyranose in a sample in which the above reagent for the
 CC determination of pyranose is reacted with pyranose in the sample and the
 CC hydrogen peroxide formed is reacted with an enzyme to develop a color, a
 CC method for diagnosing a disease accompanied by abnormality in sugar
 CC metabolism in which the above reagent for the determination of pyranose
 CC is reacted with pyranose in the sample and the hydrogen peroxide formed
 CC is reacted with an enzyme to develop a color, and a kit for pyranose
 CC analysis or the diagnosis of a disease accompanied by abnormality in
 CC sugar metabolism. The pyranose oxidase can be used as a diagnostic agent
 CC for diabetes mellitus. This sequence represents a Tricholoma matsutake
 CC pyranose oxidase protein N-terminal fragment which is described in the
 CC method of the invention.
 XX Sequence 67 AA;
 SQ Query Match 6.1%; Score 177; DB 21; Length 67;
 Best Local Similarity 57.4%; Pred. No. 6.2e-09;
 Matches 35; Conservative 6; Mismatches 20; Indels 0; Gaps 0;
 QY 409 VODGYSMPQPTFRYRPSASNVARKMADMCCEVASNLGGYLPTSPPOFMDPGLHLHLAG 468
 DB 7 VTDYGMPTFTFHKRTNADGDRDQRMNDNTNVANMLGGYLPDSYPOFMAPGLVHLHTG 66
 QY 469 T 469
 DB 67 T 67
 RESULT 12
 ID AAY81951 standard; peptide; 67 AA.
 XX AAY81951;
 AC AAY81951;
 XX 07-JUL-2000 (first entry)
 DT N-terminal fragment of Trichoderma derived antifungal protein.
 DE Antifungal protein; Pyricularia oryzae; Rhizoctonia solani; mushroom;
 KW growth inhibitor; plant pathogenic fungi; antibacterial agent;
 KW N-terminal fragment.
 XX Trichoderma matsutake.
 OS WO200014242-A1.
 PN 31-OCT-2000; 2000JP-0332085.
 PR 24-NOV-2000; 2000JP-0357102.
 PR 12-SEP-2001; 2001JP-0276832.

PD 16-MAR-2000.
 XX 19-AUG-1999; 99WO-JP04441.
 PF 08-SEP-1998; 98JP-0270606.
 PR (NISR) JAPAN TOBACCO INC.
 XX (NORQ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.
 PA Takakura Y, Kuwata S, Ohta S;
 XX WPI; 2000-256990/22.
 DR Mushroom-derived antibacterial protein against plant pathogenic fungi
 XX of rice, with activity and thermal stability, obtainable cheaply on
 PT large scale, useful in agriculture
 XX Claim 2; Page 27; 52pp; Japanese.
 PS This sequence represents an N-terminal fragment of the protein of the
 CC invention. The protein is an antibacterial protein with activity against
 CC at least Pyricularia oryzae and Rhizoctonia solani is obtained from a
 CC fraction of an aqueous extract of a mushroom precipitated by the ammonium
 CC sulphate precipitation method. The protein has a molecular weight of
 CC about 210 kD as determined by the gel filtration method, includes
 CC components of about 15 kD and 50 kD in SDS-PAGE, and is stable to
 CC heating in an aqueous neutral solution at 60 degrees C for 10 minutes but
 CC with loss of antibacterial activity after heating in the solution at 80
 CC degrees C for 10 minutes. The protein is used for inhibiting the growth
 CC of plant pathogenic fungi e.g. Pyricularia oryzae and Rhizoctonia solani.
 CC It is useful in treating rice plants, and is applicable in agriculture as
 CC an antibacterial agent. The protein has activity at relatively low
 CC concentrations, and can be produced at low cost on large scale.
 XX Sequence 67 AA;
 SQ Query Match 6.1%; Score 177; DB 21; Length 67;
 Best Local Similarity 57.4%; Pred. No. 6.2e-09;
 Matches 35; Conservative 6; Mismatches 20; Indels 0; Gaps 0;
 QY 409 VODGYSMPQPTFRYRPSASNVARKMADMCCEVASNLGGYLPTSPPOFMDPGLHLHLAG 468
 DB 7 VTDYGMPTFTFHKRTNADGDRDQRMNDNTNVANMLGGYLPDSYPOFMAPGLVHLHTG 66
 QY 469 T 469
 DB 67 T 67
 RESULT 13
 ID AAU97826 standard; protein; 539 AA.
 XX AAU97826;
 AC AAU97826;
 XX 27-AUG-2002 (first entry)
 DT Glucose dehydrogenase associated protein #2.
 DE Glucose dehydrogenase; electrode; glucose sensor; glucose-assay;
 XX substrate specificity; heat stability.
 KW Burkholderia cepacia.
 OS WO200236779-A1.
 PN 10-MAY-2002.
 XX 31-OCT-2001; 2001WO-JP09556.
 XX 31-OCT-2000; 2000JP-0332085.
 PR 24-NOV-2000; 2000JP-0357102.
 PR 12-SEP-2001; 2001JP-0276832.

XX (SODE/) SODE K.
XX Sode K;
XX WPI; 2002-463413/49.
XX N-PSDB; ABKS2367.
XX Production of Burkholderia glucose dehydrogenase for use in glucose
XX sensor electrodes and glucose-assay kits in medicine, science and
XX industry
XX
XX Claim 11; Page 56-57; 61pp; Japanese.
XX
XX The invention describes a method of producing a glucose dehydrogenase
XX comprising culturing a Burkholderia microorganism and collecting the
XX product from the medium and/or the microbial cells. Glucose dehydrogenase
XX is useful in electrodes of glucose sensors and glucose-assay kits for
XX medicine, science and industry. Glucose dehydrogenase is economically
XX produced with high substrate specificity and improved heat stability to
XX provide long-term accuracy. This is the amino acid sequence of a glucose
XX dehydrogenase associated protein described in the invention.
XX
XX Sequence 539 AA;

Query Match 6.0%; Score 173.5; DB 23; Length 539;
Best Local Similarity 19.9%; Pred. No. 4.3e-07;
Matches 109; Conservative 80; Mismatches 191; Indels 169; Gaps 24;
YY 60 PFSSAISNGKPHOREFENLSAEAVTRGVGMSTHTWCTSTPRIHP---PMESLPGIGRP 116
YY 78 PNDVYLKGEHKFNSQY-----IRAVGGTTHWAASARWFPDPMKSVYGVGR- 128
YY 117 KLSNDPAE--DKEMNELYSEAEERLIGTSTKEFDESIRHTLVLSLQDAYKDRQIP--RP 173
YY 129 ---DWPIQYDDELE--PYQRAEEELGVMPGPEE-----DLYSPRKQVPMP 171
YY 174 LPLACHR--LKNAPVEVHSAENLFHSI-----YNDKQKLLFLLTNHRCRTRIALTG 226
YY 172 LPLSFNEOTIKTA-----LNNYDPKPHVVTEPVARNSRPYDGRPTCCGNNNCWPCIPGA 226
YY 227 Y-----EKKIGAAEVRNLLATR---NPSSQL-----DSYIMAKVYVLASG 263
YY 227 MYNGIVHVEKAERAGAKLIENAVVTKETGPKRIVAAALYKDTGAHRVVEGYFVLAAN 286
YY 264 AIGNPQILYNSGFGSLQVTPRNDSLIPNLGRYITEQPMACQIVLRQEPVDSVDDPYGL 323
YY 287 GIETPKIL-----LMSANRDFPFGV 306
YY 324 PWKKEAVAOHIAKNTDALPI-----PFRDPEQVTPPTTEHPHWT-----QIH--- 368
YY 307 ANSSDMVGRNLMHDHFGTGSFVSFASEKLPWGRGPQEMTSLIGFRDGFPRATEAAKHLSN 366
YY 369 -----RDAFSYGAV--GPEVDSRV-----IVDLRFEGATDPEANNLLVFQNDVODG 412
YY 367 LSRIDQETQKIFKAGLMPDELDQAIQDRSARYVQFCFHEILFQENRIVPSKTATDA 426
YY 413 YSMPQPTFRYPSTASNVRRKMMADMCEV-----ASNIGG-----YLTPTSPQFMDP 460
YY 427 IGIPRPEITY-----AIDYVKGGAHTREVATAKVLGGTDVFNDEFAFN----- 475
YY 461 GLAHLAGTTRIGFDKATVADNLSNVDFANLYVAGNGTIRTOFGENTPLTSMCHAIS 520
YY 476 ---HITGSTINGADARDSVVDKCRTEHDENLFISSSATMPTVGTNVNLTITIALALRM 531
YY 521 ARSJINTLK 529
YY 532 S-----DTLK 536

RESULT 14
AM48440
D AM48440 standard; Protein; 1209 AA.

XX AC AAM48440;
XX DT 16-JUL-2002 (first entry)
XX DE 2-keto-D-gluconate dehydrogenase.
XX KW Cell membrane bound; 2-keto-D-gluconate dehydrogenase; enzyme;
XX KM 2,5-diketo-D-gluconate.
XX OS Unidentified.
XX PH Key Location/Qualifiers
XX FT Misc-difference 740 /note= "Encoded by TG"
XX PN KR2000019366-A.
XX PD 06-APR-2000.
XX PF 10-SEP-1998; 98XR-0037413.
XX PR 10-SEP-1998; 98XR-0037413.
XX PA (KOAD) KOREA ADV INST SCI & TECHNOLOGY.
XX PI Shin YC, Bahn JG, Yeom DY;
XX DR WPI; 2001-088033/10.
XX DR N-PSDB; ABL52917.
XX PT Base sequence of novel cell membrane-bound 2-keto-D-gluconate
XX dehydrogenase gene and method for conversion into
XX 2,5-diketo-D-gluconate - Noabstract
XX PS Claim 3; Page 9-11; 16pp; Korean.
XX CC The present invention relates to a novel cell membrane-bound
XX 2-keto-D-gluconate dehydrogenase and a method for conversion into
XX 2,5-diketo-D-gluconate. The present sequence is the protein sequence for
XX the 2-keto-D-gluconate dehydrogenase.
XX SQ Sequence 1209 AA;
Query Match 5.9%; Score 170; DB 22; Length 1209;
Best Local Similarity 20.3%; Pred. No. 3.6e-06;
Matches 109; Conservative 65; Mismatches 209; Indels 154; Gaps 21;
QY 68 NGKNPHOREFENLSAEAVTRGVGMSTHTWCTSTPRIHP---MESLPGIGRKLSDPAE 124
Db 281 NVTGNADSFQ---QGYLRTVGGTTHWAASCHRRHPSDFVMQSKYGVGR----- 327
QY 125 DDKW---NEL---YSEAEERLIGTSTKEFDESIRHTLVLSLQDAYKDRQIRIFRPLPLA 177
Db 328 ---DWPIQYDDELEPYCKAENEIGVAGP---NDPARQSPTERS-----QPYEM- 369
QY 178 CHRLKNAPEYVHSAENLFHSIYNDKQKLL-----FTLLTNHRCRTRIAL 223
Db 370 -----DMVPFAGDNYFASVYVPHGYNLVPIQGRSTRPWEGRPTCCGNNNCQIP 421
QY 224 TGGYKEKIGAAEVRNLLATRNPSQLDSYIM-----AK 256
Db 422 IGAMYNIGIHVE-----PAENGAVVLAEEAVYKMDTSDNNRITAVHMLDTSYGASHKATAK 477
QY 257 VYVLASGAIGNPQILYNSGFGSLQVTPRNDSLIPN-----LGRYITEQPMACQIVLRQEF 312
Db 478 AFALACNGIETPRLL-----MAANDANFNGIANASDMVGRNMDHSGHCSFLTKE--- 529
QY 313 VDSVRDDPYGLPWKKEAVAOHIAKNT--DALPIPFRDPEQVTPPTTEHPHWTQI---- 367
Db 530 -----PVN-----LQKGAQSSCMWGYRDGD-----PRDYSANKVILNNI 565
QY 368 -----HRDAFSYGAVGPEVDSRV-----IVDLRFEGATDPEANNLLVFQNDVODGYSM 415

```

Db      566 SRVTATQAAKKGLVGVKALDEEIRYRAVHSDLSISLEPLDPDENRLTLTKRKDPHGL 625
Qy      416 PQPTFRYRPSTASNVRAAKMADCEVASNLGGYLPTSPQOFMDPGLAL--HLAGTTTRIG 473
Db      626 PCPIIYYDVGVYVRKGAASHAQLAHEI-----GQLFDAKEFTISQGLNANNHNGGVIMG 680
Qy      474 FDKATTVADNNSLVDFANLYVAGNGTIRTOFGENPTLTSCHKAISINTLKG 530
Db      681 KNAKEAVVDGNCRAFDHENLWLPGGGAIPASVNVSTLTMAALGLKAADISLRMKG 737

```

RESULT 15

AAB97034
ID AAB97034 standard; Peptide; 34 AA.

XX AC AAB97034;

XX DT 20-JUL-2001 (first entry)

XX DE Lyophyllum shimeji antibacterial protein fragment #4.

XX KW Fungus; antibacterial; antibiotic; plant pathogen; bacterial infection;
XX KW Pyricularia oryzae; Rhizoctonia solani; rice pathogen.

XX OS Lyophyllum shimeji.

XX PN WO200121657-A1.

XX PD 29-MAR-2001.

XX XX 20-SEP-2000; 2000WO-JP06404.

XX XX 21-SEP-1999; 99JP-0267238.

XX XX (NISB) JAPAN TOBACCO INC.

XX XX (NORQ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.

XX XX Takakura Y, Kuwata S, Inoue Y;

XX XX WPI; 2001-281598/29.

XX PT Antibacterial protein and encoded gene isolated from Lyophyllum
PT shimeji, with activity against plant pathogenic bacteria, applicable in
PT agriculture e.g. rice cultivation at low concentration, produced at low
PT cost on large scale

XX PS Example 2; Page 24; 52pp; Japanese.

XX CC The present sequence is part of an antibacterial protein from the fungus
CC Lyophyllum shimeji. The protein was obtained from a fraction prepared
CC by extracting Lyophyllum shimeji with water and subjecting the extract
CC to ammonium sulphate precipitation. The protein inhibits the growth
CC of the plant pathogenic bacteria Pyricularia oryzae and Rhizoctonia
CC solani at a relatively low concentration. P. oryzae and R. solani are
CC causative of the two major diseases of rice. The protein contains
CC components of 70 kDa and 65 kDa, as determined by SDS-PAGE.
CC The antibacterial protein can be produced at low cost on a large scale.

XX SQ Sequence 34 AA;

Query Match 5.8%; Score 168; DB 22; Length 34;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 136 AERLIGTSTKEFDESIRHTLVLSLQDAYKDRQR 169

Db 1 AERLIGTSTKEFDESIRHTLVLSLQDAYKDRQR 34

Search completed: October 28, 2003, 15:57:23
Job time : 89 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2003, 15:55:56 ; Search time 29 Seconds
(without alignments)
792.234 Million cell updates/sec

Title: US-09-856-327-2_COPY_76_618

Perfect score: 2901
Sequence: 1 NABEGTAVPVVPGYHKKNEI.....IINTLKGGTDGKNTGHRNL 543

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/prodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/prodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/prodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/prodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Match	Query	DB ID	Description
1	1142.5	39.4	623	1	US-08-734-925-2
2	1119	38.6	622	3	US-09-305-381-2
3	1024.5	35.3	566	3	US-09-023-731-1
4	192	6.6	51	3	US-09-023-731-12
5	159	5.5	615	4	US-09-297-937C-9
6	159	5.5	1276	4	US-09-297-937C-13
7	130.5	4.5	58	3	US-09-023-731-11
8	115	4.0	50	3	US-09-023-731-14
9	108	3.7	35	3	US-09-023-731-15
10	105.5	3.7	778	4	US-09-198-452A-508
11	104.5	3.6	1876	2	US-08-609-049A-12
12	104.5	3.6	1876	3	US-09-170-996-12
13	102.5	3.5	61	3	US-09-023-731-6
14	102.5	3.5	1686	4	US-09-355-160D-2
15	99.5	3.4	618	4	US-09-134-001C-3169
16	96.5	3.3	1876	2	US-08-609-049A-28
17	96.5	3.3	1876	3	US-09-170-996-28
18	95.5	3.3	1257	1	US-08-049-783-2
19	95.5	3.3	1257	1	US-08-158-232-6
20	95.5	3.3	1257	1	US-08-304-626-6
21	95.5	3.3	1257	1	US-08-316-301A-6
22	95.5	3.3	1257	2	US-08-611-928-6
23	95.5	3.3	1257	3	US-09-173-891-6
24	95.5	3.3	1257	3	US-09-076-137-6
25	95.5	3.3	1257	5	PCT-US92-03624-6
26	95	3.3	3170	2	US-07-642-734C-5
27	95	3.3	3170	3	US-08-439-009A-5

28	92	3.2	585	4	US-09-352-991A-18215	Sequence 18215, A
29	92	3.2	2802	3	US-09-542-331-1	Sequence 1, Appli
30	92	3.2	2802	4	US-09-510-791-1	Sequence 1, Appli
31	91	3.1	890	4	US-09-513-783A-174	Sequence 174, App
32	90.5	3.1	806	3	US-08-549-515-5	Sequence 5, Appli
33	90.5	3.1	806	3	US-08-549-515-11	Sequence 11, Appl
34	90.5	3.1	1313	4	US-09-071-035-450	Sequence 450, App
35	90.5	3.1	1313	4	US-09-071-035-454	Sequence 454, App
36	90	3.1	487	4	US-09-198-452A-907	Sequence 907, App
37	90	3.1	543	3	US-09-199-229-2	Sequence 2, Appli
38	90	3.1	543	3	US-09-443-087-2	Sequence 2, Appli
39	90	3.1	543	4	US-09-687-298-2	Sequence 2, Appli
40	90	3.1	807	3	US-08-974-549A-5	Sequence 5, Appli
41	90	3.1	807	3	US-08-912-951-5	Sequence 5, Appli
42	90	3.1	951	4	US-09-125-635-8	Sequence 8, Appli
43	90	3.1	1132	3	US-08-851-843A-225	Sequence 225, App
44	90	3.1	1132	3	US-08-974-549A-2	Sequence 2, Appli
45	90	3.1	1132	3	US-08-974-549A-344	Sequence 344, App

ALIGNMENTS

RESULT 1
US-08-734-925-2
; Sequence 2, Application US/08734925
; Patent No. 5712139
; GENERAL INFORMATION:
; APPLICANT: NISHIMURA, IKUKO
; APPLICANT: OKADA, KIMIHARU
; APPLICANT: MINAMIHARA, TOMOYUKI
; APPLICANT: KAWAI, GENSUHIRO
; APPLICANT: KOYAMA, YASUJI
; APPLICANT: SUZUKI, MASARU
; TITLE OF INVENTION: NOVEL PYRANOSE OXIDASE, PYRANOSE OXIDASE
; TITLE OF INVENTION: GENE, NOVEL RECOMBINANT DNA AND PROCESS FOR PRODUCING
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSER: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
; ADDRESSER: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/734,925
; FILING DATE: 22-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: US 08/568,428
; FILING DATE: 06-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 7126-002-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 623 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-734-925-2

Query Match 39.4%; Score 1142.5; DB 1; Length 623;
Best Local Similarity 45.1%; Pred. No. 6.8e-112;
Matches 251; Conservative 69; Mismatches 163; Indels 73; Gaps 15;

QY 13 GYHKKNEIEFKDIDRFVNVIKGALQOVSVPRNQNVPTLDPGAWSPGSSAISNGKQP 72
DB 87 GSHKNTVEYQKNDKFNVIQGLMPVSVVNTVVDLTSPASQA--STFFVRNGANP 144
QY 73 HQREFENLSABAVTRGVGSMTHWTCSTPRIHPHMSLPGIGRPKL--SNDPAEDDKWNE 131
DB 145 EQDPLRNLGQAVTRVVGSMTHWTCATPRPK--LQRLLVKNDKADDAEWR 197
QY 132 LYSEAEERLIGTSTKFEDESIRHTLVLRSLQDAYKDRQRIFRPLPLACHRLKNAPEYVEWH 191
DB 145 EODPLRNLGQAVTRVVGSMTHWTCATPRDQ-----RPLLVKDDQADDAEWR 197
QY 198 LYTKAESYFKTGTCQKESIRHNLVNLKLAEBYKG--QRDFOQIPLAATR--RSPTFFVWS 254
QY 192 SAENLF--HSIYNDKQKLLFTLLTNHRCRLALTGGEYKIGAAEVNLLATRNPSQL 249
DB 255 SANTVFDLQNRPNTPDAPNERFNLPPAVACRY-----VRN-----TSNSEI 295
QY 250 DSY-----IMAKVYVLAGAIGNQILYNSGSL--QVTPRN--DSLIPNLGRY 295
DB 296 ESHIHDLISGRFEIKADVFVLTAGAVHNAQLLVNSGFGQLGRPDPPANPQLLSLSY 355
QY 296 ITEQPMACQIVLROEFVDSVRDQPY--GLP-----WKKEAVAQH 333
DB 356 ITEQSLVFCQVMTSLDVSXSMIIRGNFGDLGYSVYTPGAETNKHDPWNEKVKQH 415
QY 334 IAKNPTDALPIPRPEQVTPPTTEHPHMTQIHRDAFSGAVGPEVDSRVIVDLRWFG 393
DB 416 NMHQEDPLPIPFEPQVTTLPQSHPMHTQIHRDAFSGAVGQVQSIDSLIIVDMRFFG 475
QY 394 ATDEANLLVDFONDVQSGNSPQPTFRYR--PSTASVNRKMMADMCEVASNLGGVLP 452
DB 476 RTEPKENKLFSDKITDYNMPOPTDFRFPAGRTSKEAEDMTDMCVMSAKIGGFLPG 535
QY 453 SPQPMDFGLALHLAGTTRIGFDKA--TTVADNNSLVDFANLYVAGNGTIRTGGENPT 510
DB 536 SLQPMEGLVLHGGTHRMGFDEQEDKCCVNTDSRVFGKNLFLGGCGNIPYAGNPT 595
QY 511 LTSMCHAIKSARSIN 526
DB 596 LTAMSLAIKSEYIKN 611

RESULT 2

US-09-305-381-2
; Sequence 2, Application US/09305381
; Patent No. 6146865
; GENERAL INFORMATION:
; APPLICANT: Christensen, Soren
; APPLICANT: Lassen, Soren Plensted
; APPLICANT: Schneider, Palle
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides
; FILE REFERENCE: 5571-200-US
; CURRENT FILING DATE: 1999-05-05
; EARLIER FILING DATE: 1998-06-10
; EARLIER FILING DATE: 1998-06-10
; EARLIER FILING DATE: 1998-06-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 622
; TYPE: PRT
; ORGANISM: Corioliolus Versicolor
US-09-305-381-2

Query Match 38.6%; Score 1119; DB 3; Length 622;
Best Local Similarity 44.6%; Pred. No. 2.1e-109;

Matches 251; Conservative 69; Mismatches 163; Indels 60; Gaps 15;

QY 13 GYHKKNEIEFKDIDRFVNVIKGALQOVSVPRNQNVPTLDPGAWSPGSSAISNGKQP 72
DB 87 GSHKNTVEYQKNDKFNVIQGLMPVSVVNTVVDLTSPASQA--STFFVRNGANP 144
QY 73 HQREFENLSABAVTRGVGSMTHWTCSTPRIHPHMSLPGIGRPKL--SNDPAEDDKWNE 131
DB 145 EQDPLRNLGQAVTRVVGSMTHWTCATPRPK--LQRLLVKNDKADDAEWR 197
QY 132 LYSEAEERLIGTSTKFEDESIRHTLVLRSLQDAYKDRQRIFRPLPLACHRLKNAPEYVEWH 191
DB 198 LYTKAESYFKTGTCQKESIRHNLVNLKLAEBYKG--QRDFOQIPLAATR--RSPTFFVWS 254
QY 192 SAENLF--HSIYNDKQKLLFTLLTNHRCRLALTGGEYKIGAAEVNLLATRNPSQL 249
DB 255 SANTVFDLQNRPNTPDAPNERFNLPPAVACRY-----VRN-----TSNSEI 295
QY 250 DSY-----IMAKVYVLAGAIGNQILYNSGSL--QVTPRN--DSLIPNLGRY 295
DB 311 --IKAVYVILTAGAVHNAQLLAASGFGQL--GRPDPAKPLPSLLPYLGLTHITEQTLVF 364
QY 304 CQIVLROEFVDSVRDQPY--GLP-----WKKEAVAQHIAKNPTDA 341
DB 365 CQIVMSTELINSVTADMTIYKPGHPDYSVYTPGNPNKHPDMWNEKVKHMDHQEDP 424
QY 342 LPPIPRPEQVTPPTTEHPHMTQIHRDAFSGAVGPEVDSRVIVDLRWFGATDPEANN 401
DB 425 LPPIPRPEQVTPPTTEHPHMTQIHRDAFSGAVGPEVDSRVIVDLRWFGATDPEANN 484
QY 402 LLVFQNDVQDQSGNSPQPTFRYR--PSTASVNRKMMADMCEVASNLGGVLP 461
DB 485 KLFVSDKITDAYNLQPTDFR--FPGREAEADMTDMCVMSAKIGGFLPGSYQPMFPG 542
QY 462 LAHLAGTTRIGFDKA--TTVADNNSLVDFANLYVAGNGTIRTGGENPTLTSCHAIK 519
DB 543 LVLHGGTHRMGFDEKADKCCVNTDSRVFGKNLFLGGCGNIPYAGNPTLTSCHAIK 602
QY 520 SAKSIINTLKGDTGKNGTGEHRN 542
DB 603 SCEYI---KKNFEPSPNPVKHN 622

RESULT 3

US-09-023-731-1
; Sequence 1, Application US/09023731
; Patent No. 6291648
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; APPLICANT: Akihiro; Izumo, Koji.; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; TITLE OF INVENTION: GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESS: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,731
; FILING DATE: 13-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 29275/1997
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD

```

MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,731
FILING DATE: 13-FEB-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 29275/1997
FILING DATE: 13-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: KENNETH H. SONNENFELD
REGISTRATION NUMBER: 33,285
REFERENCE/DOCKET NUMBER: 3479-4000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 51
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-023-731-12

Query Match
Best Local Similarity 6.6%; Score 192; DB 3; Length 51;
Matches 32; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

QY 330 VAOHAKNPDTALPIFRDPEQVTPPTFTEHPWHQTQIHRDAFSYGAVGPE 380
Ddb 1 VEAHKKKPFDDVLPPIFFHEPEQVWIPYTSDFPWHVQVHRDAFSYGDVGPK 51

RESULT 5
US-09-297-937C-9
Sequence 9, Application US/09297937C
Patent No.: 6337199
GENERAL INFORMATION:
APPLICANT: YUM, Do Young
APPLICANT: PAN, Jae Gu
TITLE OF INVENTION: Membrane-Bound Gluconate Dehydrogenase, Gene Sequence
TITLE OF INVENTION: Encoding the Same and Production of 2-Keto-D-Gluconate
TITLE OF INVENTION: Using Transformed Recombinant E. Coli
FILE REFERENCE: P661590S0
CURRENT APPLICATION NUMBER: US/09/297,937C
CURRENT FILING DATE: 1999-05-11
PRIOR APPLICATION NUMBER: PCT/KR98/00296
PRIOR FILING DATE: 1998-09-25
PRIOR APPLICATION NUMBER: KR 97-48802
PRIOR FILING DATE: 1997-09-25
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 9
LENGTH: 615
TYPE: PRT
ORGANISM: Erwinia cyripedii
US-09-297-937C-9

Query Match
Best Local Similarity 5.5%; Score 159; DB 4; Length 615;
Matches 124; Conservative 85; Mismatches 257; Indels 158; Gaps 32;

QY 4 EGTAVPVVPGYHKKEIEBFQKIDIRFVNVI-----KGALQQVSVFVRNQNVTLPDGAWS 58
Ddb 5 ERVSVP-VSGYSRGEGVTVANELKKVDVAVVGFVGAGAIMAKELTEAGLVVALERG---- 60

QY 59 APPGSSAISNGKNP-----HOREFENLSAEAVT-----R 87
Ddb 61 --PHRDTPDGAYPQSIDELATYNIRKKLFQDLKSSTVIRHDASQTAVPYRQLAAFLPCT 118
88 GVGCGMSHTWCTSTPIHDDMESIQC-----TCBPVI SYNDEDD--VEWNEF-----VERAF 127

```

Db 119 GTGAGLHWSGVHFRVDPVELNLSHRYEARVYKGNFIEGHTIQDFGVSYNELEFPFDOAE 178
Qy 138 RLIGTSKEDFDESIRHTLVRLSLQDAY--KDRQIRFRPLP-----LACHRLKNAPEYVW 190
Db 179 KVFCTSGSAM--TIKGMIGKEKGNFYAPDRSSDF--PLPAQKRITYSAQLFAQAASVGY 235
Qy 191 H-----SAENLFPHSYND-----KOKKLP 212
Db 236 HPYDMP SANTSGPYNTYGAQMGPCNFCGVCYACVYMSKASPNVNLPALEKPEKPEL 295
Qy 213 LTNHRCRLALTGGEYKKGICA---AEVRNLLATRNPSO--LDSYIMAKVYVYLASGAIG 266
Db 296 RNNAYVLRVNLTDGKKEATGYVLDGQREV---QPADVLVLSAFQFHNHMLLSGIG 352
Qy 267 NP--QILYNSFGSLQVTPRNDLIPNL--GRYITEQPM---AFQIVLRQEFVSDVRDDPY 321
Db 353 QPVNPIITNEGVRGNFYQNI STLKALFDKNTTTPFAGAGGAVAVDDFNADNFDHGPY 412
Qy 322 GL-----PMWKEAVAQIAKNTDALTPIPRDPEQVTTPTTEHPHWHQIHRDAFSYGAV 377
Db 413 GFVGSGSPFW---VNOAGTKPVSGLPKGTNP-----WGSQWK-----AAV 450
Qy 378 GPEVDSRVIVDLRWFGATDPEANLLVFQNDVODGYSM--OPTFRYRPSSTASNVR--ARK 434
Db 451 ADTYNHHSIDAH--GAHQSYRANYLDLPYKNVYGOPLLRMTFDWQD---NDIRMAQF 505
Qy 435 MMADMCEV--ASN-----LGGYLPTSPFPQMDPGL--ALHLAGTRIGFDKATTVADNNSL 486
Db 506 MVGKMKRKITAMNPKMIIGG--AKGPGTHFTTIVYQTHMSGGAINGEDPKTSAVNRYLQ 563
Qy 487 VMDPANLYVAGNGTIRTGFCENPT 510
Db 564 SWDVPNVFVPGASAPFQGLGYNPT 587

RESULT 6

US-09-297-937C-13
; Sequence 13, Application US/09297937C
; Patent No. 6337199
; GENERAL INFORMATION:
; APPLICANT: YUM, Do Young
; APPLICANT: PAN, Jae Gu
; TITLE OF INVENTION: Membrane-Bound Gluconate Dehydrogenase, Gene Sequence
; TITLE OF INVENTION: Encoding the Same and Production of 2-Keto-D-Gluconate
; TITLE OF INVENTION: Using Transformed Recombinant E. Coli
; FILE REFERENCE: P66159US0
; CURRENT APPLICATION NUMBER: US/09/297,937C
; CURRENT FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: PCT/KR98/00296
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: KR 97-48802
; PRIOR FILING DATE: 1997-09-25
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1276
; TYPE: PRT
; ORGANISM: Erwinia cyripedii
US-09-297-937C-13

Query Match 5.58; Score 159; DB 4; Length 1276;

Best Local Similarity 19.9; Pred. No. 4.6e-07;
Matches 124; Conservative 85; Mismatches 257; Indels 158; Gaps 32;

Qy 4 EGTAVPVVPGYHKKNEIEFOKIDRFVNVJ-----KGAQQVSVPRVNRQNVPTLDPGAW 58
Db 225 ERVSVV--VSGYSGEGVTANLKKYDVAVVVGWAGAIWAKELTEGLNVVALERG--- 280
Qy 59 APGSSAISNGKVP-----HQREFNLGAEAVT-----R 87
Db 281 --PHRDTYPCGAVPOSIDELTYNIRKLFQDLKSTVTIRHDASQTAVPYRQLAFLPGT 338

Qy 88 GVGGMSTHWTCTPRHPHWPESLPG-----IGRPKLSNDPAEDD--KENWEL---YSEAE 137
Db 339 GTGGAGLHWSGVHFRVDPVELNLSHRYEARVYKGNFIEGHTIQDFGVSYNELEFPFDOAE 398
Qy 138 RLIGTSKEDFDESIRHTLVRLSLQDAY--KDRQIRFRPLP-----LACHRLKNAPEYVW 190
Db 399 KVFCTSGSAM--TIKGMIGKEKGNFYAPDRSSDF--PLPAQKRITYSAQLFAQAASVGY 455
Qy 191 H-----SAENLFPHSYND-----KOKKLP 212
Db 456 HPYDMP SANTSGPYNTYGAQMGPCNFCGVCYACVYMSKASPNVNLPALEKPEKPEL 515
Qy 213 LTNHRCRLALTGGEYKKGICA---AEVRNLLATRNPSO--LDSYIMAKVYVYLASGAIG 266
Db 516 RNNAYVLRVNLTDGKKEATGYVLDGQREV---QPADVLVLSAFQFHNHMLLSGIG 572
Qy 267 NP--QILYNSFGSLQVTPRNDLIPNL--GRYITEQPM---AFQIVLRQEFVSDVRDDPY 321
Db 573 QPVNPIITNEGVRGNFYQNI STLKALFDKNTTTPFAGAGGAVAVDDFNADNFDHGPY 632
Qy 322 GL-----PMWKEAVAQIAKNTDALTPIPRDPEQVTTPTTEHPHWHQIHRDAFSYGAV 377
Db 633 GFVGSGSPFW---VNOAGTKPVSGLPKGTNP-----WGSQWK-----AAV 670
Qy 378 GPEVDSRVIVDLRWFGATDPEANLLVFQNDVODGYSM--OPTFRYRPSSTASNVR--ARK 434
Db 671 ADTYNHHSIDAH--GAHQSYRANYLDLPYKNVYGOPLLRMTFDWQD---NDIRMAQF 725
Qy 435 MMADMCEV--ASN-----LGGYLPTSPFPQMDPGL--ALHLAGTRIGFDKATTVADNNSL 486
Db 726 MVGKMKRKITAMNPKMIIGG--AKGPGTHFTTIVYQTHMSGGAINGEDPKTSAVNRYLQ 783
Qy 487 VMDPANLYVAGNGTIRTGFCENPT 510
Db 784 SWDVPNVFVPGASAPFQGLGYNPT 807

RESULT 7

US-09-023-731-11
; Sequence 11, Application US/09023731
; Patent No. 6291648
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; APPLICANT: Akihiko; Izumo, Koji; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; TITLE OF INVENTION: GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,731
; FILING DATE: 13-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 29275/1997
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 3479-4000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792

```

: GENERAL INFORMATION:
: APPLICANT: Griffois, R.
: TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
: TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
: TITLE OF INVENTION: and treatment of infection
: FILE REFERENCE: 9710-003-999
: CURRENT APPLICATION NUMBER: US/09/198,452A
: CURRENT FILING DATE: 1998-11-24
: NUMBER OF SEQ ID NOS: 6849
: SEQ ID NO 508
: LENGTH: 778
: TYPE: PRT
: ORGANISM: Chlamydia pneumoniae
US-09-198-452A-508

```

Query Match 3.7%; Score 106.5; DB 4; Length 778;
Best Local Similarity 20.9%; Pred. No. 0.073;
Matches 84; Conservative 59; Mismatches 158; Indels 101; Gaps 20;
QY 116 PKLNDPAEDDKENWELYSERLIGTSTKEFDE-----SIRHTLVLSLQDA-----Y 164
DB 446 PRNGSPREDSPLMNALVGNHAK-HGAKTKESSSTPEISAPIVRGWSQDSVSFIV 504
QY 165 KDRQRIFRPLPLACHRLKNAPEYVWHSANLPHSIYNDKQKFLTLTNHRCRLALT 224
DB 505 MEDDHIFVDVPRKXGIDYDVPSPRWSGPARELEEDVEGD----- 543
QY 225 GYKKGKGAABVRNLLATRNPSQLDSYIMAKVYVLASGAIGNPQIILYNSGFSGLQVTPR 284
DB 544 --YEVPTSAB-----PSKDKNTYMTPR---LATPAIYD--LPSRFGSGSSRSRPS 587
QY 285 ND---SLIPNLGRVITEOPMAFCQIVLROEPVDSVRDDPYGLPWKEAVAQHIKNTPTDA 341
DB 588 SDRVRSSPN--RRGVPLPPVP--SPAMSEE--GSIYEDMSGAGAGESDYEDMSRSPS-- 640
QY 342 LPIDFRDPEQVTTPTTEHPMHTQIHRDAF-----SYGAVGPEYDSRVIVDLRMFGATDP 397
DB 641 ---PRGDLDEPIYANTPEDNPF--TORNIDRLQERSGASASVPEP-IYDEIPWIGHRPP 695
QY 398 EANNLLVFQNDVQGYCMPQ-----TFRYRSTASNVRRARQOMADMCEVASNLGGYL 450
DB 696 A-----TLPRPENTLTVNSLRVSGFGPEVRAALLSESVSAVMVEARSIV 740
QY 451 PTSP-----QFMDP--GLALHLAGTTRIGDKATTVDNNS 485
DB 741 PPTFGGGESEYLEFLGG-----LVATTKILLQKGNPRGESNA 778

RESULT 11

US-08-609-049A-12
; Sequence 12, Application US/08609049A
; Patent No. 5948664
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Molz, Lisa
; APPLICANT: Chen, Yen-Wen
; TITLE OF INVENTION: No. 5948664el PI 3-Kinase Polypeptides
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/609,049A
; FILING DATE: 29-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-063700US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1876 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-609-049A-12

Query Match 3.6%; Score 104.5; DB 2; Length 1876;
Best Local Similarity 22.0%; Pred. No. 0.55;
Matches 85; Conservative 53; Mismatches 170; Indels 79; Gaps 21;
QY 8 VPVPGYHKKNVETEFQKIDRFVNVIKGALQOVSVVNRQNVPTLDRGA-----WSAP--P 61
DB 184 VVQPAQQOORPLN--SEELQRLYSM--PAQMAVVPVQPNAYMYFGAVVTPYTAIPLV 239
QY 62 GSAISNGKNPHOREPENLSAEAVTRGVGGMSTHTCTSPRIHP--PMESLPGIG--RPKL 118
DB 240 GSAAPFPQYPAQ-----GYGFGAYTMDLRRPQSQAPQOTATTSHHSQOP 288
QY 119 SNDAEDDKENWELYSER---LIGTSTKEFDESIRHTLVLSLQDA-----YKDRQRI 170
DB 289 SNHSTSPAEANGVAFPARQVSPSTGVSSSSHTGNGHSSVPRRNDLIDLNHEDYSRV 348
QY 171 -----FRPLPLACHRLKNAPEYVWHSANLPHSIYNDKQKFLTLTNHRCRLALTG 225
DB 349 SVLEAFDPLNDNTGNDTASDSTSYAAYDPDFLYSGD-----AATQYSDPMYEAVN 401
QY 226 GYKKGKGAABVR--NLLATR-----NPSSQLDSYIMA-----KVYVLASGAIGNP-- 268
DB 402 RWDKTATVSPNVGLIGRWODFLSQSPSTSSSQYGVAPPEESLKLAEANGSGTISPPPLPP 461
QY 269 --OILYNSGFSGLQVT--PRNDSLIPNLGRVITEOPMAFCQIVLRQ-----EFVDSVR 317
DB 462 RNQCYESNOAMPVSRPQSSVLTD--SYTSSIP--ANVLDRRRTCTRLYELISDQR 516
QY 318 -DDPYGLPWKEAVAQHIKNTPTDALP 343
DB 517 TDDPELLEFY-HMVKEVRYRAPHDDAP 542

RESULT 12

US-09-170-996-12
; Sequence 12, Application US/09170996
; Patent No. 6291220
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Molz, Lisa
; APPLICANT: Chen, Yen-Wen
; TITLE OF INVENTION: No. 6291220el PI 3-Kinase Polypeptides
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/170,996
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/609,049
; FILING DATE: 29-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-063700US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:

LENGTH: 1876 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-170-996-12

Query Match 3.6%; Score 104.5; DB 3; Length 1876;
Best Local Similarity 22.0%; Pred. No. 0.55; 170; Indels 79; Gaps 21;
Matches 85; Conservative

QY 8 VPVPGHKHKNIEIFOKIDRFVNIKGAQQVSVPRNQNVPTLDPCA-----WSAP--P 61
DB 184 VPQPAQAQQORPLN-SEELQRLYSM---PAQMAVVPVQPNAYMYGAVVPTATPIV 239
QY 62 GSSAISNGKNPHQREFENLSAEAVTRGVGGMSTHTWCSTPRIHP-PMESLPGIG--RPKL 118
DB 240 GSAAFMPPOYPAQ-----GYGFGGAYTHMDLRRPQSOPAPQQTAPTTHSHSQP 288
QY 119 SNDAEDDKKEMNELYSEAE-----LIGTSKEPDESIRHTLVLSLQDA-----YKDRQRI 170
DB 289 SNHSTSSPAEANGVAPPARQVSTGVSSSHGTGNGHSSVPRRGNLDLIDLNHEDYSRV 348
QY 171 -----FRPLPLACHRLKNAPEYVEMHSAENLHFSIYNDKQKKLFTLLTNHRCRLALTG 225
DB 349 SVLEAFDPLNDNTGNDTASDTSYAEYDPDFLYSGD-----AATQYSDPMYEAVN 401
QY 226 GYEKKGIAEVR-NLLATR-----NPSQLDSYIMA-----KVYVLASGAIGNP----- 268
DB 402 RWDKTATVSPNVGLIGWRQDFLSQSFSSQYGVAPPEESLKLAEANGSOTISPPPLPP 461
QY 269 ---OILNSGFSGLQVT-PRNDSLIPNLGRYITQPMAFQOIVLRO-----EFVDSVR 317
DB 462 RNCQCVESNOAMPVSRPOSSVLTD--SYTSSIP---ANVLDRRKTCRLYELISDQR 516
QY 318 -DDPYGLPMWKEAVAQIAKNPTDALP 343
DB 517 TDDPELLEY-HMVKEVRARYPHDDAP 542

RESULT 13
US-09-023-731-6
Sequence 6, Application US/09023731
Patent No. 6291648
GENERAL INFORMATION:
APPLICANT: Kawamura, Yukio; Morita,
APPLICANT: Akihiro; Izumo, Koji.; Saka, Tomohide.
TITLE OF INVENTION: ANTI-TUMOR PROTEIN AND
TITLE OF INVENTION: GENE ENCODING SAME
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,731
FILING DATE: 13-FEB-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 29275/1997
FILING DATE: 13-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: KENNETH H. SONNENFELD
REGISTRATION NUMBER: 33,285
REFERENCE/DOCKET NUMBER: 3479-4000
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 61
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-023-731-6

Query Match 3.5%; Score 102.5; DB 3; Length 61;
Best Local Similarity 43.3%; Pred. No. 0.0023;
Matches 27; Conservative 6; Mismatches 24; Indels 5; Gaps 2;

QY 29 FVNVIKGAQQVSVPRNQNVPTLDPCAWSAPGSSA-----ISNGKNPHQREFENLSAEA 84
DB 1 FVNIINGALQPIISPSDITYOPTLAVAAW-APPIDPAEQOLVIMGNHNPQNEAGLNLP 59
QY 85 VT 86
DB 60 VT 61

RESULT 14
US-09-355-160D-2
Sequence 2, Application US/09355160D
Patent No. 6436671
GENERAL INFORMATION:
APPLICANT: Domin, Jan
TITLE OF INVENTION: No. 6436671el Lipid Kinase
FILE REFERENCE: 2332-1-004
CURRENT APPLICATION NUMBER: US/09/355,160D
CURRENT FILING DATE: 1999-10-01
PRIOR APPLICATION NUMBER: PCT/GB98/00244
PRIOR FILING DATE: 1998-01-27
PRIOR APPLICATION NUMBER: 9701652.1
PRIOR FILING DATE: 1997-01-28
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 1686
TYPE: PRT
ORGANISM: Homo sapiens
US-09-355-160D-2

Query Match 3.5%; Score 102.5; DB 4; Length 1686;
Best Local Similarity 24.3%; Pred. No. 0.74;
Matches 74; Conservative 41; Mismatches 116; Indels 73; Gaps 19;

QY 50 PTLDPGAMSAPPG-----SSAISNGKNPHQREFENLSAEAVTRGVG 90
DB 135 PTIQRGQW--PPGLPGPSTYALPSIYPTYSKQAAQFQNGFPRMPTFP--STEPYLSLP 190
QY 91 GMSTHW---TCSTPRIHPMESLPGIGRKLSDNDPAEDDKEMNELYSEAEAL-----ICT 142
DB 191 GSPYFSYPLTPATP-PH-PQGSLLP-IYRPVVDMA---KLFDKIASTSEFLKNGKART 244
QY 143 STKEPDESIRHTLVLSLQDAYKQRQIRFRL-----PLACHRLKNAPEYVEMHSAENLPHS 199
DB 245 DLEITDSKVSNLQVSPKSEDISK-----FDWLDDPLSKPKVDNV-EVLDEHEEKVSSL 298
QY 200 IYNDKQKKLFTLLTNHRCRLALTGGVYKKGIAEAVENLNLATRNPPSQLDSYIMAKYVV 259
DB 299 LAKDPWAVLLEERSTANC-----HLERKVNKGSUSVATVTRSQSLNIRTOLAK--- 348
QY 260 LASGAIGNPQILYNSGFSGLQVTPRNDSLIPNLGRYITEQPMAFQOIV--LRQBFVDSVR 317
DB 349 -AQGHISQKD---PNGTSSL---PTGSSLQLEV-EVQNEENAAFCRSITKLTKEF----- 395
QY 318 DDPY 321
DB 396 --PY 397

Search completed: October 28, 2003, 16:01:11
Job time : 32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2003, 15:59:57 ; Search time 72 Seconds
(without alignments)

1262.932 Million cell updates/sec

Title: US-09-856-327-2_copy_76_618

Perfect score: 2901

Sequence: 1 NAEAGTAVYVGVYHKNEI.....IINTLKGDTGKNTGHRML 543

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 629382 seqs, 167460630 residues

Total number of hits satisfying chosen parameters: 629382

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/FCR_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/FCRUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	1024.5	35.3	566	10	US-09-911-969-1
2	192	6.6	51	10	US-09-911-969-12
3	130.5	4.5	58	10	US-09-911-969-11
4	129	4.4	545	15	US-10-156-761-9468
5	115	4.0	50	10	US-09-911-969-14
6	108	3.7	35	10	US-09-911-969-15
7	106.5	3.7	775	11	US-09-920-843A-58
8	106	3.7	546	15	US-10-253-971-3
9	104	3.6	514	15	US-10-156-761-9160
10	103.5	3.6	342	15	US-10-156-761-14017
11	102.5	3.5	61	10	US-09-911-969-6
12	102.5	3.5	1686	14	US-10-092-219-2
13	102	3.5	1244	9	US-09-845-915-8
14	102	3.5	1244	12	US-10-393-316-8
15	97.5	3.4	1536	10	US-09-801-368-322

16	97	3.3	804	15	US-10-128-714-3088
17	97	3.3	1028	15	US-10-128-714-8088
18	96	3.3	970	15	US-10-026-021-2
19	95.5	3.3	1257	9	US-09-738-363-6
20	95	3.3	1778	12	US-10-238-075-749
21	94	3.2	1286	15	US-10-156-761-14323
22	94	3.2	1714	15	US-10-128-714-3176
23	94	3.2	1750	15	US-10-128-714-8176
24	92	3.2	733	12	US-10-291-583-60
25	92	3.2	1381	12	US-10-032-585-7784
26	91.5	3.2	299	10	US-09-764-864-1522
27	91.5	3.2	350	10	US-09-764-864-1102
28	91.5	3.2	757	15	US-10-014-789A-4
29	91	3.1	568	10	US-09-873-880-24
30	91	3.1	1390	15	US-10-100-957A-174
31	91	3.1	847	15	US-10-097-340-35
32	90.5	3.1	847	15	US-10-177-293-498
33	90	3.1	492	10	US-09-881-752A-142
34	90	3.1	807	15	US-10-044-692-5
35	90	3.1	951	12	US-10-044-539-5
36	90	3.1	1132	10	US-10-379-616-8
37	90	3.1	1132	10	US-09-990-080-2
38	90	3.1	1132	10	US-09-749-728B-31
39	90	3.1	1132	10	US-09-843-676-225
40	90	3.1	1132	10	US-09-953-052-2
41	90	3.1	1132	12	US-10-295-681-57
42	90	3.1	1132	15	US-10-053-758-225
43	90	3.1	1132	15	US-10-208-243-2
44	90	3.1	1132	15	US-10-054-295-225
45	90	3.1	1132	15	US-10-054-611-225

ALIGNMENTS

RESULT 1

US-09-911-969-1
; Sequence 1, Application US/09911969
; Patent No. US20020137896A1
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; Akihiro; Izumo, Koji.; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/911.969
; FILING DATE: 24-Jul-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/023.731
; FILING DATE: 13-FEB-1998
; APPLICATION NUMBER: JP 29275/1997
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 3479-4000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792

Sequence 3088, Ap
Sequence 8088, Ap
Sequence 2, Appli
Sequence 6, Appli
Sequence 749, App
Sequence 14323, A
Sequence 3176, Ap
Sequence 8176, Ap
Sequence 60, Appl
Sequence 7784, Ap
Sequence 1522, Ap
Sequence 1102, Ap
Sequence 4, Appli
Sequence 24, Appl
Sequence 174, Appl
Sequence 35, Appl
Sequence 498, App
Sequence 142, App
Sequence 5, Appli
Sequence 8, Appli
Sequence 2, Appli
Sequence 31, Appl
Sequence 225, App
Sequence 2, Appli
Sequence 57, Appl
Sequence 225, App
Sequence 2, Appli
Sequence 225, App
Sequence 225, App

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 566
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-911-969-1

Query Match 35.3%; Score 1024.5; DB 10; Length 566;
Best Local Similarity 43.0%; Pred. No. 3.4e-90;
Matches 232; Conservative 77; Mismatches 180; Indels 51; Gaps 12;
QY 2 AEEGTAVYVPGYHKKNIEFQKIDRFVNIKGLQVSVPRNQNVPTLDPCAWAPP 61
DB 62 AEIGSQDNPIVGAHKNSIKFKQIDKFVNIINGALQISPSDSTQPTLVAAM-APP 120
QY 62 GSSA---ISGNKPHOREFENLSAEAVTRGVGGMSTHTWCTSTRIHPMESLFGIGRPK 117
DB 121 IDPREGQVIMGNPNQAGLNLPQSAVTRIVGGNATHWTCACTPHD-----EE 170
QY 118 LSNPAEDDKENNELYSABERLIGTSTKEFDESIRHTLVLSLQDAYKDRQRIERPLPLA 177
DB 171 RVNPNV-DKQEFDALEAKTLNHNHSDQYDSDIRQIVVKETLQOTL-DASRGVTTILPG 228
QY 178 CHRLKNAPEYWEHSAENLFHSIYNDKQKFLFTLLTNHRCRTRALTGCGYEKKIGAAEVR 237
DB 229 VERRTDNPIYVTTGAD---TVLGDVPKSPRFALVTETRTVKLIVSETNPTQVVAALLR 284
QY 238 NLLATRNPSQDLSVIMAKVVLASGAIGNQILYNSGFGSLQVTPRNDLSLIPNLGRYIT 297
DB 285 NL-----NTSN--DELIVAKSVIAGAVCTQILWNS-----NIRPY-----ALGRYLS 327
QY 298 EQPMAFQVILVRQFVDSVRDDPYGLPMWKEAVAQIAKNTDLPPIPRDPEPQVTTTF 357
DB 328 EQSMTFCQVILKRGIVDAIATDPR-----FAAKEVAHKKKHDDVLPPIPFHEPEQVWIPY 383
QY 358 TEEHPWHTQIHRDAFSYGAQVDSRVIVDLRFMGATDPANNLLVF-----QND 408
DB 384 TSDPFMHWQVHRDAFSYGDVGPKADPRVVDLRFEGKSDIVEENRVTFGNPKLREWEAG 443
QY 409 VODGVSMPQPTFRYRPSFASVNRARKMADNCEVASNLGGLYPTSPQFMDPGLALHLAG 468
DB 444 VTDYGMQPTFHVKRTNADGDRQNMNDNTVNMGLGGYLPQSGYPOFMAPGLVHLHTG 503
QY 469 TTRIGFDKATTVADNNSLVMDFAMLYVAGNCTIRTFGENPTLTSMCHAIKSARSIIINTL 528
DB 504 TTRIGTDQTSVADPTSKVHNFNLLWVGNGCIPDATACTNTRTSVAYALKGAEVNVYL 563

RESULT 2
US-09-911-969-12
Sequence 12, Application US/09911969
Patent No. US20020137896A1

GENERAL INFORMATION:
APPLICANT: Kawamura, Yukio; Morita,
Akhihiro; Izumo, Koji.; Saka, Tomohide.
TITLE OF INVENTION: ANTITUMOR PROTEIN AND
GENE ENCODING SAME

NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD 97
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/911,969
FILING DATE: 24-Jul-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/023,731
FILING DATE: 13-FEB-1998
APPLICATION NUMBER: JP 29275/1997
FILING DATE: 13-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: KENNETH H. SONNENFELD
REGISTRATION NUMBER: 33,285
REFERENCE/DOCKET NUMBER: 3479-4000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 751-6849
TELEFAX: (212) 751-6849
TELEX: 421792

INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 51
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-911-969-12

Query Match 6.6%; Score 192; DB 10; Length 51;
Best Local Similarity 62.7%; Pred. No. 4e-11;
Matches 32; Conservative 6; Mismatches 13; Indels 0; Gaps 0;
QY 330 VAQHIKNTDLPALPIPRDPEQVTPFTTEHPWHTQIHRDAFSYGAQVPE 380
DB 1 VEAHKKHDPDVLPIPFHEPEQVMTPTSDPFPHVQVHRDAFSYGDVGPK 51

RESULT 3

US-09-911-969-11
Sequence 11, Application US/09911969
Patent No. US20020137896A1

GENERAL INFORMATION:
APPLICANT: Kawamura, Yukio; Morita,
Akhihiro; Izumo, Koji.; Saka, Tomohide.
TITLE OF INVENTION: ANTITUMOR PROTEIN AND
GENE ENCODING SAME

NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/911,969
FILING DATE: 24-Jul-2001

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/023,731
FILING DATE: 13-FEB-1998
APPLICATION NUMBER: JP 29275/1997
FILING DATE: 13-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: KENNETH H. SONNENFELD
REGISTRATION NUMBER: 33,285
REFERENCE/DOCKET NUMBER: 3479-4000

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 751-6849
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:

Db 337 VSKEPNDTKFKQTLALHDWYLGSDDDWYPLGGIOMLGKSDAEQIHGEAPRWAGAPDMP 396
QY 383 SRVI-----VDLWFQATDPEANLLVFNQVDGYSMPQPTFRYPSTASNVARKOMAD 438
Db 397 FEVLAAHVAVDFWLCGEDPLAENRVTLDDGGIHLALDE-----KNNTAGLKRRLRHLQ 451
QY 439 MCEVASNLGG-----YLPTSPQPMDFMDPLALHLAGTTRIGDFDKATTVDNNSLVW 488
Db 452 M---LHLMGHEHLLSHSYLHKMP-----IGATAHQAGTVRFGRDPADSDALDVNCKAH 504
QY 489 DFNALYVAGNGTIRTFGENPTLTSMCHAISARSIINTLK 529
Db 505 DLDNLVYVDTSFPFGVAVNPSTLTAIALNALRVGDHLAERLR 545

RESULT 5

US-09-911-969-14
; Sequence 14, Application US/09911969
; Patent No. US20020137896A1
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; Akihiro; Izumo, Koji.; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/911,969
; FILING DATE: 24-Jul-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/023,731
; FILING DATE: 13-FEB-1998
; APPLICATION NUMBER: JP 29275/1997
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 3479-4000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-09-911-969-14

Query Match 4.0%; Score 115; DB 10; Length 50;
Best Local Similarity 53.5%; Pred. No. 0.0011;
Matches 23; Conservative 5; Mismatches 15; Indels 0; Gaps 0;
QY 409 VQDGYSPQPTFRYPSTASNVARKWADWCEVASNLGGYLP 451
Db 8 VTDYGMFPQPTFHVKRTNADGDRDQRMNDMTNVANMLGGYLP 50

RESULT 6

US-09-911-969-15

LENGTH: 58
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-911-969-11
Query Match 4.5%; Score 130.5; DB 10; Length 58;
Best Local Similarity 44.4%; Pred. No. 4.6e-05;
Matches 28; Conservative 11; Mismatches 13; Indels 11; Gaps 2;
QY 258 YVLASGAINGPQILYNSGFGSLQVTPRNDLIPNLGRYTEQPMARFCQIVLRQEFVDSVR 317
Db 2 FVIACGAVCTPQILWNS-----NIRPV-----ALGRYLSEQSMTEFCQIVLRKGIIVDATA 50

QY 318 DDP 320

Db 51 TDP 53

RESULT 4

US-10-156-761-9468
Sequence 9468, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 9468
LENGTH: 545
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-9468

Query Match 4.4%; Score 129; DB 15; Length 545;
Best Local Similarity 19.2%; Pred. No. 0.0023;
Matches 100; Conservative 57; Mismatches 188; Indels 176; Gaps 19;
QY 126 DKWNNELYSEARLIGTSTK-----EPDESIRH-----TLVLSIQDAYK 165
Db 84 DKHGNQFPPEVNYVGNTKFVGALFRLRPEDFGE-LRHHDGISPAWFLSYEELEPYT 142
QY 166 DRQIF-----RPLPLACHRLKNAPYEW-HSAE-----NLPH-----SIY 201
Db 143 QAEHLVLVHGRHGDEPTGPTSAQYAPVQHEPRIQQLSHDLEKQGLHPHLPVGNLT 202
QY 202 NDDROKKLFTLLTNHRCRLALTGTGYEKKGA-----A-EVRN 238
Db 203 QDDRGRATHA-----SACIRCDRVDFGFCVLGSAKSAQVCDVPALEHANVEMLTADYVR 258
QY 239 L-----LATRNPSSQLDSYIMAKVVVVLASGAINGPQILYNSGFGSLQVTPRND 286
Db 259 LDTATGTSVTSVATVGDGASTVEFSADIVVACGAVNSAVILLRSA-----307
QY 287 SLIPNLGRYTEQPMARFCQIVLRQEFVDSVRDD--PYGLPMWKEAVAQHIKNPTDALPI 344
Db 308 -----DDRHPQGLANSSGVGVGRHYMRHNNLALMA 336
QY 345 PFRPEQVTTFTTEHPWH-----TQHRDAFSY-GAVGPEVD 382

Sequence 15, Application US/09911969
Patent No. US20020137896A1
GENERAL INFORMATION:
APPLICANT: Kawamura, Yukio; Morita, Akihiro; Izumo, Koji.; Saka, Tomohide.
TITLE OF INVENTION: ANTITUMOR PROTEIN AND GENE ENCODING SAME
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/911,969
FILING DATE: 24-Jul-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/023,731
FILING DATE: 13-FEB-1998
APPLICATION NUMBER: JP 29275/1997
FILING DATE: 13-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: KENNETH H. SONNENFELD
REGISTRATION NUMBER: 33,285
REFERENCE/DOCKET NUMBER: 3479-4000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 35
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-09-911-969-15
Query Match 3.7%; Score 108; DB 10; Length 35;
Best Local Similarity 66.7%; Pred. No. 0.0031;
Matches 22; Conservative 2; Mismatches 9; Indels 0; Gaps 0;
QY 453 SPPOFMDPLGLHLAGTTRIGFDKATTVADNNS 485
DB 2 SYPOFMAGLVHLHTGTTRIGTDDQTSVADPTS 34
RESULT 7
US-09-820-843A-58
Sequence 58, Application US/09820843A
Publication No. US20030039963A1
GENERAL INFORMATION:
APPLICANT: Council of Scientific and Industrial Research
TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES
FILE REFERENCE: Q63915
CURRENT APPLICATION NUMBER: US/09/820,843A
CURRENT FILING DATE: 2001-03-30
NUMBER OF SEQ ID NOS: 118
SOFTWARE: PatentIn version 3.0
SEQ ID NO 58
LENGTH: 775
TYPE: PRT
ORGANISM: C. pneumoniae CWL029
FEATURE:
NAME/KEY: misc_feature

OTHER INFORMATION: hypothetical protein
NAME/KEY: misc_feature
OTHER INFORMATION: gi|4376756
US-09-820-843A-58
Query Match 3.7%; Score 106.5; DB 11; Length 775;
Best Local Similarity 20.9%; Pred. No. 0.62;
Matches 84; Conservative 59; Mismatches 158; Indels 101; Gaps 20;
QY 116 PKLSNDPAEDDKENNELYSEAEELIGTSTKEPDE-----SIRHTVLVRLQDA-----Y 164
DB 443 PRNGSPREDSPLMNALVGVAAHK-HGAKTKESSESSTPEISISAPIVRGWSQSSVSFIV 501
QY 165 KDRQRIREFPLPLACHRLKNAPEYVEVHSAENLPHSYINDDKQKKLFTLLTNHRCRLALT 224
DB 502 MEDDHIFVDVPRKDGIVDPSSPRWSPARELEEDVFGD----- 540
QY 225 GGYEKKIGAAEVRNLLATRNPSQSDSYIMAKYVIVLASGAIGNPQILYNSGFGSLQVTPR 284
DB 541 --YEVPI TSAE-----PSKDKNIYMTPR---LATPAIYD--LPSRFGSGSSRSRPS 584
QY 285 ND---SLIENLGRYTEQPMFQCQIVLROEFVDSVRDDPYGLPFWKEAQAHTAKNPTDA 341
DB 585 SDRVRSSSPN--RGVPLPPVP--SPAMSEE--GSIYEDMSGAGAGESDYEDMSRSPS-- 637
QY 342 LPIPRDPEFQVTPPTTEHPWHTQIHRDAF-----SYGAVGPPEVDSRVIVDLRMFGATDP 397
DB 638 ---PRGDLDEPIYANTPDNPP-TQKNIDRILOERSGGASASVPEP-IYDEIPIWHGRPP 692
QY 398 EANNLLVFQNDVQDGYSMFOP-----TFVRVPSTASNVRRKMMADMCEVASNLGGYL 450
DB 693 A-----TLRPENTLTNLSLRVSPGFGPEVRAALLSESVAMVVEAESIV 737
QY 451 PTSPP-----QFMDP--GLALHLAGTTRIGFDKATTVADNNS 485
DB 738 PPTPEPGDGESEYLEPLGG---LVATTKILLQKQWPRGESNA 775
RESULT 8
US-10-253-971-3
Sequence 3, Application US/10253971
Publication No. US20030070192A1
GENERAL INFORMATION:
APPLICANT: GEORGES, FAWZY
APPLICANT: DONG, JIN-ZHUO
APPLICANT: KELLER, WILF
APPLICANT: HOUSAIN, ATTA A. K.
APPLICANT: SELVARAJ, GOPALAN
APPLICANT: DATLA, RAJU
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING SECONDARY PLANT
TITLE OF INVENTION: METABOLITES
FILE REFERENCE: 73776-159D
CURRENT APPLICATION NUMBER: US/10/253,971
CURRENT FILING DATE: 2002-09-25
PRIOR APPLICATION NUMBER: US 60/072156
PRIOR FILING DATE: 1998-01-22
PRIOR APPLICATION NUMBER: US 09/012453
PRIOR FILING DATE: 1998-01-23
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 546
TYPE: PRT
ORGANISM: Arthrobacter pascens
US-10-253-971-3
Query Match 3.7%; Score 106; DB 15; Length 546;
Best Local Similarity 20.6%; Pred. No. 0.39;
Matches 72; Conservative 45; Mismatches 131; Indels 102; Gaps 16;
QY 210 FTLLTNHRCRLALTGGYEKKIGAAEVRNLL--ATRNPSQSDSYIMAKYVIVLASGAIGN 267
DB 224 FTLLTGLRARQLVFDA--DKRCTGVDVDSAFGRTHLSARCE-----VILSTGAIDS 274

QY 268 POILYNSGFS-----GLQVTRNDLSLIPNLGRYTEQPMARCOIVLRQEFVDSVRDD 319
DB 275 PKLLMLSGIGAPAAHAEHGVLEVIV--DS--PGVGEHLQDHEGVVQPEAKQOMVQT----- 326
QY 320 PYGLPMWKEAV-----AQHAKNPTDALPIPRDP-----EPOVTPPT 358
DB 327 --STQWWEIGIFTPTEGLDRPDLMMHGVSPFDMVTLRGVPTTENGFSLTPTNVT----- 380
QY 359 BEHPWHTQIHRDAFSYQAV---GPEVDSRVIVDLRWFGATDPEANNLLVFQNDVQDGYSM 415
DB 381 -----HARSRGTVLRSDRFRDKPAVDPRYE--TDPEGHDMRVVMVAGIR----- 422
QY 416 POPTFRYRPSTASNRARKMADN-----CEVASNLGGVLPSPQFMDPGLALH 465
DB 423 -----KAREIAQAPMAEWGTRELSFGTEAQTDELDQYIRKT-----HNTVYH 466
QY 466 LAGTTTRIG-FDKATTVADNNSLVDFANLYVAGNGTIRTGFGENPTLTSM 514
DB 467 PVGTVMGPADDDMSPLDPELRVKGVTGLRVADASVMPHEVTVNPNITVM 516

RESULT 9
US-10-156-761-9160
; Sequence 9160, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9160
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9160

Query Match 3.6%; Score 104; DB 15; Length 514;
Best Local Similarity 20.3%; Pred. No. 0.56;
Matches 92; Conservative 52; Mismatches 183; Indels 126; Gaps 20;
Y 152 RHTVLRSLQDAYKDRQRIFRPLPLACHRLKNAPYVEWHIS-----AENLPHSIYNDKQ 206
b 125 RHRLQNRIVPVADKDRNFIARDFVTAASALGPVPIEDFNAPFAAGAGFFSLAYEPESN 184
Y 207 KK-----LFTLLTNHRCRTRALALGGY-----EKKIGAAEVRNLLATRNPSQLDSVI 253
b 185 RRSASVAYLHPVLDPRNLTRLEWTAYGLLPDGGRLTRVQROSDGT-----TATVR 238
Y 254 MAKVYVLASGAIGNPQILYNSGFSGLQ-----VTPRNDSLIPNLGRYTEQPMARCOIV 307
b 239 AARDMLLCAGAITDPRLLLSGVGPAQLRDLGIEVRAD--VPGVGENLLDHPE-----VI 293
Y 308 LRQEF-----VDSVRDDDPYGLRWKEAQAQHNKPTDALPIPRDPQV-----TTPFT 358
b 294 VMEHTGGLPPNSAMDSADGL-----FLRRAP-----RGRPDLMFHFYQVPT 336
Y 359 EE-----HPWHTQIHRDAFSYG-AVGPEVDSRVIVDLRWFGATDPEANNLLVFQNDVQD 412
b 337 VNTERLGHF-----AIHGVCMTPNVPARSVGRMVLADADPTR----- 375

QY 413 YSMPTFRYRPSTASNRARKMADNCEVASNLGGVLP-----TSPQFMDPGL 462
DB 376 -----QPCLODFRYETDPGYDEKTIIVDGLGIAREVAAAAPLURDLKREVAPEGPIRTDAEL 431
QY 463 A-----LHACTTRIGFPDKATTVADNNSLVDFANLYVAGNGTIR-----TGGEN 508
DB 432 SAVGRKAANTVHPAGTCRMG-----ADDDPLAVLDPQLRVGVSGVRVWDASLFPTM 484
QY 509 PTLTSMCHAIKARSIIINTLKGGTGDKNTGEHR 541
DB 485 PTINPMLTVLLAAERAVDLIKG---TRQPGEAR 514

RESULT 10
US-10-156-761-14017
; Sequence 14017, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14017
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14017

Query Match 3.6%; Score 103.5; DB 15; Length 342;
Best Local Similarity 23.7%; Pred. No. 0.33;
Matches 89; Conservative 41; Mismatches 109; Indels 137; Gaps 23;
QY 191 HSAENL-----FHSIYNDKQKXLFLLTNHRCRTRALALGGYKKGKGAAB-VRNLLAT 242
DB 40 HAVESLGTARAGGMHLAYVTNNALRTPDAVADH-LTELGIPTGPSDVITSAQAVARLISE 98
QY 243 RNPSSQLDSYIMAKVYVLASGAIGNPQILYNSGFSGLQVTRNDLSLIPNLGRYTEQPM 302
DB 99 QVPAG-----ARVLVI-----GGEGLRVALRERGLEP----- 125
QY 303 FCQIVLRQEFVDSVRDDP-----YG---LPMWKEAQAQ-HIAK-----NPTDALPIP- 345
DB 126 -----VDSAEEDDPAVVQGVGGPELPMGRFAEASAIARGVPWFASNTD-LTIPS 174
QY 346 -----FRDEPQVT-TPFTEHPWHTQIHRDA-FSYGAVGCP-EVDSRV 385
DB 175 ARGAPNGAAVEWRIATGAEPQVAGKPL-----PPMRETILRTGAERPLVVGDR 227
QY 386 IVDLRWFGATDPEANNLLVFQNDVQDGYSMFPQTFRFRPSTASNRARKMADNCEVASN 445
DB 228 DTDLE--GAFNGEVDLSLLV--TGVTGQAQLLAAPRHRPT-----YVDAD 269
QY 446 LGGYLPSTPQFMDPGLALHLAGTTRIGFDKATTVADNNSLVDFANLYVAGNGTIRTG 505
DB 270 LRGLL-TQPEWBEAGDGRCGWT-----ATAGSE-----RLELAGEAMDG- 312
QY 506 GENPTLTSCHAIKSA 521
DB 313 -----LRALCAAATA 323

RESULT 11

US-09-911-969-6
; Sequence 6, Application US/09911969
; Patent No. US20020137896A1
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; Akahiro; Izumo, Koji.; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/911.969
FILING DATE: 24-Jul-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/023,731
FILING DATE: 13-FEB-1998
APPLICATION NUMBER: JP 29275/1997
FILING DATE: 13-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: KENNETH H. SONNENFELD
REGISTRATION NUMBER: 33,285
REFERENCE/DOCKET NUMBER: 3479-4000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 61
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-09-911-969-6

Query Match 3.5%; Score 102.5; DB 10; Length 61;
Best Local Similarity 43.5%; Pred. No. 0.026;
Matches 27; Conservative 6; Mismatches 24; Indels 5; Gaps 2;
QY 29 FNVVIGALQVSVFVRNQVPTLDPGAWSPGSSA-----ISNGKNPHOREPENSAAEA 84
Db 1 FVNTINGALQVSVFVRNQVPTLDPGAWSPGSSA-----ISNGKNPHOREPENSAAEA 84
QY 85 VT 86
Db 60 VT 61

RESULT 12

US-10-092-219-2
; Sequence 2, Application US/10092219
; Publication No. US20020115114A1
; GENERAL INFORMATION:
; APPLICANT: Domin, Jan
; TITLE OF INVENTION: No. US20020115114A1el Lipid Kinase
; FILE REFERENCE: 1064HG/50947
; CURRENT APPLICATION NUMBER: US/10/092,219
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: PCT/GB98/00244
; PRIOR FILING DATE: 1998-01-27
; PRIOR APPLICATION NUMBER: 09/355,160

; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: 9701652.1
; PRIOR FILING DATE: 1997-01-28
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 2
; LENGTH: 1686
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-219-2

Query Match 3.5%; Score 102.5; DB 14; Length 1686;
Best Local Similarity 24.3%; Pred. No. 5.2; Indels 73; Gaps 19;
Matches 74; Conservative 41; Mismatches 41; Indels 73; Gaps 19;
QY 50 PTLDPGAWSPGSSA-----SSAISNGKNPHOREPENSAAVTRGVG 90
Db 135 PTIQRCQW--PGLPGPSTYALPSIYPTSYSKAAAFQNGFNPRMPTFP--STPIYLSLP 190
QY 91 GMSTHW-----TCSTPRIHPMESLPGIGRPKLSNDPAEDDKENNELYSEARL-----IGT 142
Db 191 GQSPYFSLPTATP-PH-PQGSLLP-IYRPVSTDMA---KLPDKIASTSEFLKNGKART 244
QY 143 STKEFDESIRHTLVLSRLQDAYKDRORIFRPL---PLACHRLKNAEYVWHSAENLFHS 199
Db 245 DLEITDSKVSNLQVSPKSEDISK-----FDWLDLDELSPKVDNV-EVLDEHEEKVSS 298
QY 200 IYNDKQKKLFTLLTNHRCRLALTGGEYKKGICAAEVRNLLATRNPSQLDSYIMAKVYV 259
Db 299 LAKDPWDVLLERSTANC-----HLERKVNKSLSVATVTRSSQSLNRTTQLAK--- 348
QY 260 LASGAIGNPQILYNSGFGSLQVTPRNDLSLIPNLGRYITEQPMACQIV--LRQEFVDSVR 317
Db 349 -AQGHISQKD---PNCSTSL---PTGSSLLQEV-EVQNEEMAAFCRSITKLKTKF----- 395
QY 318 DDY 321
Db 396 --PY 397

RESULT 13

US-09-815-915-8
; Sequence 8, Application US/09815915
; Patent No. US20020025931A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 3714, 16742, 23546, AND 13887 NOVEL
; TITLE OF INVENTION: PROTEIN KINASE MOLECULES AND USES THEREFOR
; FILE REFERENCE: 38155-20006.00
; CURRENT APPLICATION NUMBER: US/09/815,915
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 60/191,846
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-815-915-8

Query Match 3.5%; Score 102; DB 9; Length 1244;
Best Local Similarity 17.4%; Pred. No. 3.6; Indels 190; Gaps 23;
Matches 95; Conservative 65; Mismatches 65; Indels 190; Gaps 23;
QY 34 KGALQVSVFVRNQVPTLDPGAWSPGSSAISNGKNPHOREP-----ENL 80
Db 661 EGPLTAITIP-----RPSVASTQSTSGSPHCQPEKDLQPMPTVELYSRENPF 711
QY 81 SBAVTRG---VCGMSTHWTCTSTPRIHPNMESLPGIGRPKLSNDPAEDD-----KE 128
Db 712 SGLVVTGEPPSGSRTDLQLQIDHIGHDN--LPNRESNKSQDLGPKELPDHNLVRE 769

QY 129 WNELYSEA-ERLIGTSTKEFDESI---RHTLVLSLOD---AVKDRQIRFRPLACHR 180
Db 770 FENLPGTEESKILLESDEDEKLSRGQHCIEISLPGDLVIVEKDHSAATEPLDV---825
QY 181 LKNAPEVVEHSAENLFHSIYNDKOKKLTLLTNHRCRTRALTTGGYEKKIGAAEVRNLL 240
Db 826 ---TKTQTFSVVPNDKNEIMKLLT---VGTSEI---854
QY 241 ATRNPSQLDSYI---MAKYVVLASGAIGNPQILYNSGFSGLQ---VT 282
Db 855 ---SRDIDPHVEGGQGVQVAEMQKNKISKDDIMSEDLFGHGDLSLTFHQBKREKIT 910
QY 283 PRNDSL---IPNLGRYITE---298
Db 911 PRNGELFHCVSENEHGAPTRKDMVRSSFVTRHSRIPVLAQEIDSTLESSSPVSAKEKLLQ 970
QY 299 ---QPMACQIVLRQEFVDSVRDDPYGLPMWKEVAQAHIKNTDALTPIPRDRPQVPT 354
Db 971 KKAYQPDVLKLVKQKFSFLGD---LSSASDKLLEKLTATVPAPF---CEEVL 1020
QY 355 TPFT---BEHPWHTQIHRDAPSYGAVGPEVDSRVIVDLRWFGATDPEANLLVFQNDVQ 410
Db 1021 TFSRLTVDSH---LSSAEDSFLSPIISQSRKSKIPRPVSWNTDQVNSSTSQF---1073
QY 411 DGYSMPQPTFRYRPSTASNVRAR---KMDMCEVASNLGGYLPSPQ 456
Db 1074 ---FPRPPPG---KPTRPGVEARLRRYKVLGSSNSDSDFSLAQILON-GSQKPRSTTQ 1127
QY 457 FMDPG 461
Db 1128 CKSPG 1132

RESULT 14

US-10-393-316-8
Sequence 8, Application US/10393316
Publication No. US20030175786A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: Meyers, Rachel
TITLE OF INVENTION: 3714, 16742, 23546, AND 13887 NOVEL
FILE REFERENCE: PROTEIN KINASE MOLECULES AND USES THEREFOR
CURRENT APPLICATION NUMBER: US/10/393,316
CURRENT FILING DATE: 2003-03-20
PRIOR APPLICATION NUMBER: US 09/815,915
PRIOR FILING DATE: 2001-03-23
PRIOR APPLICATION NUMBER: US 60/191,846
PRIOR FILING DATE: 2000-03-24
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8
LENGTH: 1244
TYPE: PRT
ORGANISM: Homo sapiens
US-10-393-316-8

Query Match 3.5%; Score 102; DB 12; Length 1244;
Best Local Similarity 17.4%; Pred. No. 3.6;
Matches 95; Conservative 65; Mismatches 195; Indels 190; Gaps 23;
Y 34 KGAQQSVVVRNQNVPLDQWAPPGSSAISKGNPHQREF-----ENL 80
b 661 EGPLTAITIP-----RPSVASTQSTSGFHCQKQPEKDLQPMETVELYSPRENF 711
Y 81 SAEAVTRG---VGGNSTHTWCTSPRIHPMESLPGIGRPKLSNDPAEDD-----KE 128
b 712 SGLVVTGEPSPGSGSRDLQIDQIGHDM---LPNIRESNKSQDLGPKLEPDHNLVVRE 769
Y 129 WNELYSEA-ERLIGTSTKEFDESI---RHTLVLSLOD---AVKDRQIRFRPLACHR 180
b 770 FENLPGTEESKILLESDEDEKLSRGQHCIEISLPGDLVIVEKDHSAATEPLDV---825

QY 181 LKNAPEVVEHSAENLFHSIYNDKOKKLTLLTNHRCRTRALTTGGYEKKIGAAEVRNLL 240
Db 826 ---TKTQTFSVVPNDKNEIMKLLT---VGTSEI---854
QY 241 ATRNPSQLDSYI---MAKYVVLASGAIGNPQILYNSGFSGLQ---VT 282
Db 855 ---SRDIDPHVEGGQGVQVAEMQKNKISKDDIMSEDLFGHGDLSLTFHQBKREKIT 910
QY 283 PRNDSL---IPNLGRYITE---298
Db 911 PRNGELFHCVSENEHGAPTRKDMVRSSFVTRHSRIPVLAQEIDSTLESSSPVSAKEKLLQ 970
QY 299 ---QPMACQIVLRQEFVDSVRDDPYGLPMWKEVAQAHIKNTDALTPIPRDRPQVPT 354
Db 971 KKAYQPDVLKLVKQKFSFLGD---LSSASDKLLEKLTATVPAPF---CEEVL 1020
QY 355 TPFT---BEHPWHTQIHRDAPSYGAVGPEVDSRVIVDLRWFGATDPEANLLVFQNDVQ 410
Db 1021 TFSRLTVDSH---LSSAEDSFLSPIISQSRKSKIPRPVSWNTDQVNSSTSQF---1073
QY 411 DGYSMPQPTFRYRPSTASNVRAR---KMDMCEVASNLGGYLPSPQ 456
Db 1074 ---FPRPPPG---KPTRPGVEARLRRYKVLGSSNSDSDFSLAQILON-GSQKPRSTTQ 1127
QY 457 FMDPG 461
Db 1128 CKSPG 1132

RESULT 15

US-09-801-368-322
Sequence 322, Application US/09801368
Patent No. US20020128250A1
GENERAL INFORMATION:
APPLICANT: Busby, Robert
APPLICANT: Call, Brian
APPLICANT: Hecht, Peter
APPLICANT: Holtzman, Doug
APPLICANT: Madden, Kevin
APPLICANT: Maxon, Mary
APPLICANT: Milne, Todd
APPLICANT: No. US20020128250A1man, Thea
APPLICANT: Royer, John
APPLICANT: Salama, Sofie
APPLICANT: Sherman, Amir
APPLICANT: Silva, Jeff
APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
FILE REFERENCE: 109272.147
CURRENT APPLICATION NUMBER: US/09/801.368
CURRENT FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: US 09/487,558
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: US 60/160,587
PRIOR FILING DATE: 1999-10-20
NUMBER OF SEQ ID NOS: 440
SOFTWARE: PatentIn version 3.0
SEQ ID NO 322
LENGTH: 1536
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-801-368-322

Query Match 3.4%; Score 97.5; DB 10; Length 1536;
Best Local Similarity 19.4%; Pred. No. 14;
Matches 103; Conservative 71; Mismatches 163; Indels 193; Gaps 28;
QY 58 SAPGSSAISKGNPHQREFNELSA-----EAVTRGVGGNSTHTWCTSPRIHPMESLP 111
Db 1044 SGSDGSGSASRKPKYQOEMSLLDILHRSYQKUKR-----SNDEKGVPOLSPEEP 1098
QY 112 G-IGRPKLSNDPAEDD-----KEMNELYSEA 136

```
Db      1099 NTIEEEELIDBEAKNPWLTGNLVEANSQGIIONRSIFNLPANTNVIYFPRHWTIIY--- 1155
Qy      137  ERLIGTSTKEPDESIRHTLVRLSL-----QDAYKDRQIRFRP 173
Db      1156 ERL--EIKQNNERTKINTRESTVTFPAKLDLSSQLSEMGDFVGEDAYKQVRLSR- 1212
Qy      174  LPLACHRLKNAPEYVEVHSAENLFHSIYNDKQKLFTL-----LTNHRCTRUALTGGY 227
Db      1213 -----RLINGDLEHWF--EESLRQAYN-NKAFKLYTIDKVTQSLVKHAHT--LMTDAK 1261
Qy      228  EKKIGAAEVRNLLATRNPSSQLDSYIMAKVYVVLASGAIGNPQIILYNSGFSGLQVTPRND 287
Db      1262 TAEIMALFVKD-----RN-----ASTTSAKDQIIYR-----LQVR-SHMS 1295
Qy      288  LIPNLGR-----YITEQPMACQIVLROEFVDSVRDDPYGLPWNKEAVAQHIKNPT 339
Db      1296 NTENMFRIEFKRTILHVSIOYIALDDLTLKEPKADEK-----WKYVYTSYALPHPT 1347
Qy      340  DALPIPPRDPFPQVTPTEHPHWTQIHRDAFSYGAVGPEVDSRVIVDLRMFGATDPEA 399
Db      1348 EGI-----PHEKLKIPLE-----RLIEFGQDIDGTEVDEEF-----SPEG 1383
Qy      400 >NNL-----LVFQNDVQDQ-YSM--PQTPPYRPSSTASNVPRKMMADMCEVASNLG 447
Db      1384 ISVSTLKIKIQPITYQLHIENGSDYVFTRKATNKY-PTIANDNTQKGWVQKKELIS--- 1439
Qy      448  GYLTPSPQFMD--PGLALHLAGTTRIGFDK-----ATTVADNNSL 486
Db      1440 -----KFLDCAVGLNLDLDEAKLSMQKMKWENLKDSIAKTSAGNQGI 1481
```

Search completed: October 28, 2003, 16:09:44

Job time : 74 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run on: October 28, 2003, 15:54:42 ; Search time 42 Seconds
(without alignments)
1243.324 Million cell updates/sec

Title: US-09-856-327-2_COPY_76_618
Perfect score: 2901
Sequence: 1 NAEETAVPVPGVHKQNEI.....IINTLKGTGDKNTGERNL 543

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR 76:.*
1: PIR1:.*
2: PIR2:.*
3: PIR3:.*
4: PIR4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Match	Length	DB ID	Description
1	170.5	5.9	573	E81385	probable oxidoredu
2	166.5	5.7	579	H87451	oxidoreductase, GM
3	159	5.5	615	B38575	glucuronate 2-dehydr
4	158.5	5.5	529	A84260	hypothetical prote
5	158.5	5.5	619	AB2003	dehydrogenase chal
6	144	5.0	722	B75608	GMC oxidoreductase
7	141	4.9	591	H83362	glucuronate dehydr
8	140.5	4.8	562	JC7628	glucoside 3-dehydr
9	135	4.7	561	AE3094	oxidoreductase Atu
10	135	4.7	561	E98192	probable oxidoredu
11	128.5	4.4	748	T10651	hypothetical prote
12	122	4.2	527	AD3082	dehydrogenase Atu4
13	122	4.2	527	E98202	dehydrogenase chal
14	116	4.0	502	T18562	hypothetical prote
15	116	4.0	514	B70917	probable zwf2 prot
16	116	4.0	599	T19711	hypothetical prote
17	114	3.9	1012	T00958	hypothetical prote
18	113.5	3.9	578	F70736	probable choD prot
19	111.5	3.8	494	AD3255	l-sorbose dehydr
20	110.5	3.8	589	T50698	probable mandelon
21	110	3.8	1498	AF1082	B. subtilis Yuka p
22	108	3.7	1125	T19193	hypothetical prote
23	106.5	3.7	775	B72074	hypothetical prote
24	106.5	3.7	775	C81594	hypothetical prote
25	106.5	3.7	775	D86549	hypothetical prote
26	105.5	3.6	1611	T38236	hypothetical prote
27	105	3.6	499	T45749	hypothetical prote
28	104	3.6	1917	C88728	protein C48A7.1 li
29	102.5	3.5	768	JC6564	cellobiose oxidase

ALIGNMENTS

RESULT 1

E81385
Probable oxidoreductase chain Cj0415 [imported] - Campylobacter jejuni (strain NCTC 1116)
C/Species: Campylobacter jejuni
C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C/Accession: E81385
R/Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillin:
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell
Nature 403, 685-688, 2000
A/Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A/Reference number: A81250; MUID:20150912; PMID:10688204
A/Accession: E81385
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-573 <PAR>
A/Cross-references: GB:AL139075; GB:AL111168; NID:G6967817; PIDN:CAB74251.1; PID:G696788:
A/Experimental source: serotype O2, strain NCTC 11168
C/Genetics:
A/Genes: Cj0415

Query Match	5.9%	Score 170.5;	DB 2;	Length 573;
Best Local Similarity	20.9%	Pred. No. 4.9e-05;		
Matches 111;	Conservative 61;	Mismatches 237;	Indels 123;	Gaps 21;
QY	77	FENLSAEAVTRGVGGMSTHTWCTSPRIHPMSLPGIRP-----KLSNDPAEDD-----K	127	
Db	84	YRMGSGFLIGNNVGGAGVHMGWTFPMFPYDFEIQTLKQRYGNKLGNDYTLQDWGVYK	143	
QY	128	EWNELYSEAEIRLIGTSTKEF-----DEIRHTLVLSLQDAYKDRQIRFR	172	
Db	144	DMEPYDREKTCGVSGEFPNLAEMKAGFRSPYPOEPLNTMLKRFESAASSNLHTY	203	
QY	173	PLPLA-----CHRLKVAPEYVHSAENLFHSIYNDKOKKLF	210	
Db	204	RLPANSKGGYTPDQDLAPCQYCAICERFG-----CEYGAKASPLNTVIPKAMSTGKY	258	
QY	211	TLLTNHRCRLALTGGYKKGICAAEVNLLATRNSSQLDSYIM-AKVTVLASGAGNQ	269	
Db	259	TIRTSYNTVQIL-----KDGKVTGVKFDVTRT-----MKEYIQPADIVLVTSYMFNNAK	308	
QY	270	ILYNSGFSGLQVTP-----RNDSLIPNLG--RYITEQPNACQIVLRQEFVDSVRDD	319	
Db	309	LMVSNL-GEQVDPKTGKTGLGRNYCYQNMGTATFAFDEQFNTF-----MGSGALGTSSD	363	
QY	320	PYG--LPWKE-----AVACHIAKNPTDALPIPRPEPQVTPFTTEEHFWHTQIHRD	370	
Db	364	FNGDNFDSKKEFLGAMTYSVQLGTRPIQSPAPAGAP-----TWGAEF-KK	410	
QY	371	APSYG-----AVGPEVDSRVIVDLRWFGATPEANLLVFQNDVQDGYSMPOPTFRYRPS	425	
Db	411	ALNVNFTRAITVGGQ-----GASLPHKNYLSLDPTYKDAFGMPLRLITYN-F	457	

Db 564 SWDVNVFVFGASAPFQGLGNPT 587

RESULT 4

AB4260
hypothetical protein Vng1035c [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: AB4260
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabil
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: AB4160; MUID:20504483; PMID:11016950
A:Accession: AB4260
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-529 <80>
A:Cross-references: GB:AE004437; NID:g10580587; PIDN:AAG19445.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG1035C

Query Match 5.5%; Score 158.5; DB 2; Length 529;
Best Local Similarity 22.3%; Pred. No. 0.00036;
Matches 117; Conservative 44; Mismatches 194; Indels 169; Gaps 22;
QY 84 AVTRGVGGMSTHTWCTSTPRIHP---MESLPGIGRPKLSNDPAEDDKENNELYSEERLI 140
b 88 ARVKGVGSTLHWQNMWMLHEQDFLASATGVA-----DWPFVDTLKPYAAAESAL 142
Y 141 -----GSTKEFDESIRHTLVRLSLQDQYKQRIRFRLPLACHRLKNAPEYVEMHSAENLF 197
b 143 GYSGASDNFPAPPREPQHPAPPPSYSD---SLFAD---ACESLGIAT----- 185
Y 198 HSIYNDKQKLLFTLLTNHRCRLALTG-----GYE-----KKIG 232
b 186 HSPVN-----ARLSAGRETRACVGTGCPVCPGSAKYDVAHVDRATDAGARVID 237
Y 233 AAEVRNL-----LATRNPSSQLDSYIMAKVYVLASGAIGNPQIL-----Y 272
b 238 EAPVQLEHDAAGDRVTGAVATPDGTHROS---ATEFVLAAAGLETPELLLSDSRY 294
Y 273 NSGF---SGLQVTPRNDLIPNLGRYITQPMAFCQIVLRQBFVDSVRDDPYGLPWKEA 329
b 295 PDGLANSSGL-----VGRY-----FMD----- 311
Y 330 VAQHTAKNPTDALPIFRDPPEQVTPPTTEHPHWTQIHRDAFSGAV-----GPEV 381
b 312 ---HLFAGAGGTLDEFTQNHVGFNT---TESQYV---DRPDGSRGAIKLEFLNAGPSP 363
Y 382 DSRVVDLRWFCA-----TQPEANLLVFQNDVQDGYSMPOPTFR 421
b 364 REMALSGDDWGDAMCDRIDASGTHIAUGLVEQPRPENRVLHPERTDVHGNVPDVV 423
Y 422 YRPSTASNVRAKMMADMCEVASNLGGVLP-TSPQFMDPGLALHAGTTRIGFDKATTV 480
b 424 WLSAYERRTIERANEIQREILLTELGADEMTVGPE--DTGPAFHHMGTTRMGTDPAESV 481
Y 481 ADNNSLVNDFANLYVAGNGTIRTFGENPTLTSMCHAIKSARSI 524
b 482 VDPRLRTHLSNLSVASSSVFPTAGAMNPTLTIAALKAADHI 525

RESULT 5

B2003
hydrogenase chain [imported] - Noctoc sp. (strain PCC 7120)
C:Species: Noctoc sp. PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
A:Accession: AB2003

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anai
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AB2003
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-619 <60>
A:Cross-references: GB:BA000019; PIDN:BA077942.1; PID:g17135396; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr1576

Query Match 5.5%; Score 158.5; DB 2; Length 619;
Best Local Similarity 19.7%; Pred. No. 0.00046;
Matches 125; Conservative 72; Mismatches 217; Indels 221; Gaps 27;
QY 22 FQKIDRFVNVIKALQOVSVPRNQNVPTLDPGANSAPPGSSA-----INSKKN 71
Db 60 FOSYVDTFYKAVD-----KNPNSP-----YFANSNVQSPDYNDYFIEQGPM 101
QY 72 PHOREPENLSASAVTRGVGGMSTHTWCTSTPRIHP---PMESLPGIGRPKLSNDPAEDDKKE 128
Db 102 P-----LAGSYTRVLGGTTMHWEAKTPRLPEDFKLSSTYGGQL-----DWPIDYHD 148
QY 129 MNELYSEAEERLIGTSTK-----EPDESIRHTLVRLSLQDQYKQRIRFRLPLACHRLKNAPEYVEMHSAENLF 170
Db 149 LEFYRKAEHENGVCDDVDEQALGLEFPQD---YFPMKELPPSYLDQKIEKVGNTNVE 206
QY 171 -----FRLPLACHRLKNAPEYVEMHSAENLF-----HSIYNDK----- 205
Db 207 LYGKTHLTSFSTFPOARNGVPN-PKY---DOGNLFVPDGVTSVHPVQYGERCOGNANCV 261
QY 206 -----QKKLFTLLTNHRCRLALTGCGYKKGKIGAAEVRNLLATRNPSQLD 250
Db 262 PICPQVQKYDARTLSKAFETGKQHVLYQVAVKVEYDQTRITAIHYKHYKKNSEY 321
QY 251 SYIMAK---VYVLASGAIGNPQILYNSGSGLOVTPRNDLIPNLGRYITEQPMACQIVL 308
Db 322 TTGIAGKTLFVLATNAVENARLLGSDL-----PNTSLI-----GYL----- 360
QY 309 ROEFVDSVRDDPYGLPWKEAVAQHIKKNPTDALPIFRDPE-----PQVTPPF----- 357
Db 361 -----MDHFTLAW-----ALM-----PEVTGTMRGPLVTSGIGTFR 392
QY 358 -----TESHPHWTQIHRDAFSGAVGVPEDSVRVID-----LR 390
Db 393 KGDFRKKQSAFVADLHNDGWMATGSPKSEVEDAVDNKNKYQELRQTLISIRQLLLA 452
QY 391 WFGATDPPEANLLVFQNDVQDGYSMPOPTFRYRPSASNV---RARKMMADMCEVA---S 444
Db 453 FMCELLPEYGNRVITDPRHKKLG-----NVRPVINENLPDYSRRTLAYTKVSRVIFE 506
QY 445 NLGG-----YLPSTPQPMDFGLAL-----HLACTTRIGFDKATTVADNNSLVNDFANL 493
Db 507 RLGAEDYTHYDQDPAYEPFEGEGYVYKGNHFSGTHIMGTTPLSNVVDSYLSRWDHKNL 566
QY 494 YVAGNGTIRTFGENPTLTSMCHAIKSARSIINTL 528
Db 567 FLVGAGSMPTIGSSNTLTIAALSFTAEHMLQEL 601

RESULT 6

B75608
GMC oxidoreductase - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
A:Accession: B75608
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; I M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.M. Science 286, 1571-1577, 1999

Best Local Similarity 23.7%, Pred. No. 0.0097;
Matches 45; Conservative 29; Mismatches 77; Indels 39; Gaps 5;
QY 372 PSY-GAVGPEVDSRVIVDLR-----W-----FGATDEANNLLVFQNDVQ 410
Db 379 FGVOGAASREGWDREIAELNIGADLKQALTOPGGWTIGTGMFLPDHNRISLDHSVR 438
QY 411 DGYSMPQPTFRYRSTASNRVRKRWADMC-----VASNLGGVLPSPQFWDPG 461
Db 439 DKWGLPVLSDVELKQNERDVRDMVDVADLLEAAGVKNVKGVDGYA-----PG 489
QY 462 LAHLAAGTTRIGFDKATTVADNNSLVWDFANLYVAGNGTIRTGFGENTLTSMCHAKSA 521
Db 490 MGIHENGTMGRDPTKTSVLNSHQVWDAPNVFTDGCACHTSSSCVNPSTLYMALTARAV 549
QY 522 RSIINTLKGG 531
Db 550 DYAVEELKRG 559
RESULT 9
AE3094
Oxidoreductase Atu4377 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AE3094
A:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan,
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AE3094
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-561 <KUR>
A:Cross-references: GB:AE008699; PIN:AA14517.1; PID:g17742847; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
A:Genetics:
A:Gene: Atu4377
A:Map position: linear chromosome
Query Match 4.7%; Score 135; DB 2; Length 561;
Best Local Similarity 24.8%; Pred. No. 0.026;
Matches 39; Conservative 25; Mismatches 59; Indels 34; Gaps 3;
Y 392 FGATDEANNLLVFQNDVQGYSMQPTFRYRSTASNRVRKRWADMC----- 441
b 420 FGEMLPYHDNRVKLDHDKKDWGL-----PVLSSNMVEMKQNELDMEDMVDVAVEMF 471
Y 442 -----VASNLGGVLPSPQFMDPGLAHLAGTTRIGFDKATTVADNNSLVWDFANLY 494
b 472 EAVGINKVNPSPRGTYA-----PGMGIHENGTMGRDPTKTSVLNGNQQVWDAPNVF 522
Y 495 VAGNTGIRTFGENPTLTSMCHAKSAIINTLKGG 531
b 523 VTDGACMTSASCNPSTLYMALTARAAEFASVSRKKG 559
RESULT 10
E98192
Probable oxidoreductase chain C0415 [imported] - Agrobacterium tumefaciens (strain C58,
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: E98192
A:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Woliam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: E98192

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-561 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK89063.1; PID:g15158859; GSPDB:GN00170
C:Genetics:
A:Gene: AGR L 980
A:Map position: linear chromosome
Query Match 4.7%; Score 135; DB 2; Length 561;
Best Local Similarity 24.8%; Pred. No. 0.026;
Matches 39; Conservative 25; Mismatches 59; Indels 34; Gaps 3;
QY 392 FGATDEANNLLVFQNDVQGYSMQPTFRYRSTASNRVRKRWADMC----- 441
Db 420 FGEMLPYHDNRVKLDHDKKDWGL-----PVLSSNMVEMKQNELDMEDMVDVAVEMF 471
QY 442 -----VASNLGGVLPSPQFMDPGLAHLAGTTRIGFDKATTVADNNSLVWDFANLY 494
Db 472 EAVGINKVNPSPRGTYA-----PGMGIHENGTMGRDPTKTSVLNGNQQVWDAPNVF 522
QY 495 VAGNTGIRTFGENPTLTSMCHAKSAIINTLKGG 531
Db 523 VTDGACMTSASCNPSTLYMALTARAAEFASVSRKKG 559
RESULT 11
T10651
Hypothetical protein TSP17.20 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Sep-2000
C:Accession: T10651
A:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancroft
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z16533
A:Accession: T10651
A:Molecule type: DNA
A:Residues: 1-748 <BEV>
A:Cross-references: EMBL:AL049917; GSPDB:GN00062; ATSP:T5F17.20
A:Experimental source: cultivar Columbia; BAC clone T5F17
C:Genetics:
A:Gene: ATSP-T5F17.20
A:Map position: 4
A:Introns: 70/3
C:Superfamily: Arabidopsis thaliana hypothetical protein T5K18.160
Query Match 4.4%; Score 128.5; DB 2; Length 748;
Best Local Similarity 19.7%; Pred. No. 0.13;
Matches 105; Conservative 77; Mismatches 187; Indels 165; Gaps 24;
QY 76 EPNLSABAVTRGVGCMSTHWTCTSPRIHPPMESLPGIGRPKLSNDPAEDDKEMNELYSE 135
Db 301 KFMLLAGSAVG---GGTAVNWSAS---ITPDHVL-----QEW-----SE 334
QY 136 AERLIGTKSF-----DE-SIRHTLVLSLDQAYKDRORIPRPLPLACHRL-----K 182
Db 335 GSKIFFGQYQSAWDEVTIRIGVTCVKGFGON-----QVLKGCERLGLQVESVPR 389
QY 183 NAPE-----YVEMHSAENLPHSIYNDKQKLLTLLTHRCRRLAL-----TGYEKK 230
Db 390 NSPDHYCGLCGYCGRAGAKNGTDQTLWLDVAVENGAVILGTGKAERFVLVDNTSSNERK 449
QY 231 -----IGAAEVRNLIATRNPSQLDSYIMAKVYVLASGAIGNPQILYNSGSLQVTPRN 285
Db 450 KRCVGFVASSVGGKIGKK-----FIEARVTSSAGSLTTPMLSSGLKN----- 495
QY 286 DSLIPNLR-----YITEQMAFCQIVLRQEFVDSVR--DDPYGLPMMWEAVA 331
Db 496 ---PNIGRNKLHPVLMTCWYFPEKSEFGKMGYEGGIITSVHMND-----TESGC 544
QY 332 QHIAKNPTDALPIPRDEPOVTPPTTEHPWHQIHRDAFSYAGVGEVDSRVIVDLRW 391
Db 545 KAILENPLIG-PASYAGLSPPWVS-----GPDKERMIK----- 576

QY 392 FGATDPEANLLVFQNDVQDYSM--PQTFRRPSTASNVRAKWMADCEVAS---NL 446
Db 577 YGRT-----AHLFALVRDLGSGEVNMENEVYRTTKDRENLRAGLRQALRSVAAGAVEV 632
QY 447 GGY-----LPTSPQFMDPGLAL-----HLAGTTRIGFDKAT 478
Db 633 GTYRSDGQKKCEAITKEMEEFLDEVDVAGGVGTGKEYWYTFSAHOMGSCRMGVTABE 692
QY 479 TVADNNSLVWDFANLYVAGNGTIRGTGENTPTLTSMCHAIKSAKSIINTLKGGT 532
Db 693 GALDENGESNEAELGFLVCDGSIPLSAVGVPNMITIQSTAYCISSKIVDSIQNK 746

RESULT 12
AD3084
dehydrogenase Atu4296 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AD3084
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AD3084
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-527 <KUR>
A:Cross-references: GB:AE08689; PIDN:AA45090.1; PID:gl7742757; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu4296
A:Map position: linear chromosome

Query Match 4.2%; Score 122; DB 2; Length 527;
Best Local Similarity 19.8%; Pred. No. 0.23;
Matches 101; Conservative 55; Mismatches 175; Indels 180; Gaps 22;
QY 75 REPENLAEAVTRGVGGMSTHTCTSTPRIHP---PMESLPGIGRPKLSNDPAEDDKWNE 131
Db 104 RDFREVEHE-----DGVSPAMPVSVAELBFFYAEARLFGV-RGRAGDDPTPPRSAPY 156
QY 132 LYSEA--ERLIGTSTKEFDESIRHTLVLSQDAYKDRQIRPLPLA-----CHRL 181
Db 157 MHADIPHEPVIGRVAKGFER-----LGLRPPH-----MPSAIDYGGGLCRR 199
QY 182 KNAPEYVEVHSAENLFSYNDKQKLP-----TLTNHRCRTRALTGGYEKKIG 232
Db 200 GTCDAFVCRFDAG-----DAETRLRLPALRHPNVSLTGARVRL-IAODGDKHIV 250
QY 233 AAEVRNLLATRNPSQLSDSYIMAKVYVLASAGNPOILYNSGSGQLVTPRNDSLIPN- 291
Db 251 AVEIER-----AGEITT-IEAPLFVLSAGAINSALILRSA-----DEKPNG 292
QY 292 -----LGRYITEQPM-----FCQIVLROEFVDSVRDDP----- 320
Db 293 LANSVGVRGLMNHLSGLMGLLPFTINDTRFPKTSLSLNDFFDGTGDEAARGNVQMLG 352
QY 321 -----YGLPWWKEAQAQIAKNPTDALPIPRDPEPQVTTFTTEHPWHTQIHRDAF 372
Db 353 NIQGPIMIRAAYPMPRPPLANLLARHSVDFLW-----SED----- 387
QY 373 SYGAVGPEVDSRVIVDLRFWFGATDPEANLLVFQNDVQDYSMPQPTFRYRPSSTASN- 429
Db 388 -----TPKYDSRV-----KPMGKNGAE-----LIYRPGDREAHQR 417
QY 430 -VRARKWADMCCEVASNLGGYLPSPQFMDPGLALHAGTTRIGFDKATTVADNNSLVW 488
Db 418 FVRHRSLLRKNKGFPPVVLGHSFGIEAPS-----HOCGTVMRGDDPKKAALNALCQTY 469

RESULT 14
T18562
hypothetical protein F15C11.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T18562; T20970
R:Barlow, K.

QY 489 DFANLYVAGNGTIRGTGENTPTLTSMCHAIK 519
Db 470 DHPNLYVVDAGFPFSSAALNPALTVAQAALR 500

RESULT 13
E98202
dehydrogenase chain I (AF068066) [imported] - Agrobacterium tumefaciens (strain C58, Ce
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: E98202
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman
A: Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tu
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: E98202
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-527 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK89143.1; PID:gl5158953; GSPDB:GN00170
C:Genetics:
A:Gene: AGR L1138
A:Map position: linear chromosome

Query Match 4.2%; Score 122; DB 2; Length 527;
Best Local Similarity 19.8%; Pred. No. 0.23;
Matches 101; Conservative 55; Mismatches 175; Indels 180; Gaps 22;
QY 75 REPENLAEAVTRGVGGMSTHTCTSTPRIHP---PMESLPGIGRPKLSNDPAEDDKWNE 131
Db 104 RDFREVEHE-----DGVSPAMPVSVAELBFFYAEARLFGV-RGRAGDDPTPPRSAPY 156
QY 132 LYSEA--ERLIGTSTKEFDESIRHTLVLSQDAYKDRQIRPLPLA-----CHRL 181
Db 157 MHADIPHEPVIGRVAKGFER-----LGLRPPH-----MPSAIDYGGGLCRR 199
QY 182 KNAPEYVEVHSAENLFSYNDKQKLP-----TLTNHRCRTRALTGGYEKKIG 232
Db 200 GTCDAFVCRFDAG-----DAETRLRLPALRHPNVSLTGARVRL-IAODGDKHIV 250
QY 233 AAEVRNLLATRNPSQLSDSYIMAKVYVLASAGNPOILYNSGSGQLVTPRNDSLIPN- 291
Db 251 AVEIER-----AGEITT-IEAPLFVLSAGAINSALILRSA-----DEKPNG 292
QY 292 -----LGRYITEQPM-----FCQIVLROEFVDSVRDDP----- 320
Db 293 LANSVGVRGLMNHLSGLMGLLPFTINDTRFPKTSLSLNDFFDGTGDEAARGNVQMLG 352
QY 321 -----YGLPWWKEAQAQIAKNPTDALPIPRDPEPQVTTFTTEHPWHTQIHRDAF 372
Db 353 NIQGPIMIRAAYPMPRPPLANLLARHSVDFLW-----SED----- 387
QY 373 SYGAVGPEVDSRVIVDLRFWFGATDPEANLLVFQNDVQDYSMPQPTFRYRPSSTASN- 429
Db 388 -----TPKYDSRV-----KPMGKNGAE-----LIYRPGDREAHQR 417
QY 430 -VRARKWADMCCEVASNLGGYLPSPQFMDPGLALHAGTTRIGFDKATTVADNNSLVW 488
Db 418 FVRHRSLLRKNKGFPPVVLGHSFGIEAPS-----HOCGTVMRGDDPKKAALNALCQTY 469

RESULT 14
T18562
hypothetical protein F15C11.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T18562; T20970
R:Barlow, K.

submitted to the EMBL Data Library, August 1997

A;Accession: Z18976

A;Status: preliminary; translated from GB/EMBL/DBD

A;Molecule type: DNA

A;Residues: 1-502 <WIL>

A;Cross-references: EMBL:Z98262; PIDN:CAB10932.1; GSPDB:GN00019; CESP:F15C11.2

A;Experimental source: clone VF15C11L

R;Wilkinson, J.

submitted to the EMBL Data Library, April 1996

A;Accession: Z19352

A;Status: preliminary; translated from GB/EMBL/DDEU

A;Molecule type: DNA

A;Residues: 1-502 <WIL>

A;Cross-references: EMBL:Z71260; PIDN:CAA95799.1; GSPDB:GN00019; CESP:F15C11.2

A;Experimental source: clone F15C11

C;Genetics:

A;Map position: 1

A;Introns: 31/3; 67/1; 169/3; 212/3; 293/2; 294/1; 340/3; 368/3; 406/2; 428/2; 448/3; 48

Query Match

Best Local Similarity 4.0%; Score 116; DB 2; Length 502;

Matches 110; Conservative 65; Mismatches 201; Indels 126; Gaps 26;

40 VSVVRNQNVPVLDGAWSPGSSAISNGKPHQREFENLSAEVTRGVGGMSTHTWCS 99

74 VHLVIRNQARPTPAPAA-ATPTASSAPSS--NP----- 103

100 TPRIHP-----PMESLPGLIGRP-KLSNDP-----ABDDKEMNELYSEAR--RLIGTSTKE 146

104 TPSSQPNTPNPFPAAGMGSGPADILNPDAMRSVNDNITQOLGNPEFMTITQSNPQ 163

147 F-----DESIRHTL-----VLSLDQAYKDR---QRIFRPLPLACHRLKNAPEYVEVHSA 193

164 FOALIERPNEVGHIILNDPNVMQTMEMIRPNMFOEMMRNHDOAIRNLOGIP-----GG 217

194 ENLFHSIYNDKQKFLTLTNHRCITRLALTG-----GYEKKIGAAEVRLNLLATR 243

218 EAALERLYN-DVQEPFLNSATN-----SLSNPTFASLGDQSSFRVDRAGCENNEALP 270

244 NPSSQDSYINAKVYVVLASGAIIGNPQILYNSGFGSLQVTPRNDLSLIPNLGRYITQPMAP 303

271 NPWAS-----NANQATNQSNNRSADFNLLSDSGISSLEQMSNPSQASMF 319

304 COIVLRQEFVDSRDDPYGLPWWKEAVAQHIANKP--TDAL--PIPRFRDPEPQVTTTPE 359

320 -----SPEVINSIR-----QNMSSNPLDLSIVGQIPSDARDNPQISEGIRR 360

360 EHPWHTQTHRDAPSYGAV-GPEVDSRVIVDLRWFGATDPEANNLL-VFQNDVDQGYSMQP 417

361 SFPQMLNMWSDPSVMEAMRNPRVSEAFRQIQEGFSTLRREAPQLNLNFOAGAMGGGAFGS 420

418 PTERYRPSSTASNRARKKMDMCEVASNLGGVLPSPQFMDP--GLALHLGATTRIGF- 474

421 DA-----NASSAGANSAGLADLFN-SMNMGGRPSSTAAPVNPEQTYASQLEQLQSMGFS 475

475 DKATTVADNNSLVMDFANLYVA 496

476 DRARNA---ALTATFGDLNAA 494

RESULT 15

70917

robable zwf2 protein - Mycobacterium tuberculosis (strain H37RV)

Species: Mycobacterium tuberculosis

Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

Accession: B70917

Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

Connor, R.; Davies, R.; Feltwell, T.; Dentles, S.; Hamlin, N.; Holroyd, S.

ajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

ature 393, 537-544, 1998

A;Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome :
A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: B70917

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-514 <COL>

A;Cross-references: GB:Z95844; GB:AL123456; NID:g3250713; PIDN:CAB09259.1; PID:g2131049

A;Experimental source: strain H37RV

C;Genetics:

A;Gene: zwf2

C;Superfamily: glucose-6-phosphate dehydrogenase

Query Match

Best Local Similarity 4.0%; Score 116; DB 2; Length 514;

Matches 98; Conservative 48; Mismatches 163; Indels 120; Gaps 23;

29 FVNVIKALQQ-VSVVVRNQNVPVLDGAWSPGSSAISNGKPHQREFENLSAEVTR 87

77 FQOVYNAVOEHCRTPFRQONWDRLAEG-FRVPGTFDDDDAFQAETLEKLDAB---R 132

88 GVGGMSTHTWCTSTPRIHPPM-ESL--PGIGRPKLSNDPAEDDKEMNELYSEAERLIGTST 144

133 GTGNGHAFYLAIPPKSFPVVCQLHKSLGARFQ-----GDRMSRVVIE----- 175

145 KEFDESIRHTLVLSLDQAYKDRQIRFRLPLACHRL--KNAPEYVEVHSAENLFHSIYN 202

176 KPFQGHDLASARELNKAVNAVPEEAVFR-----IDHYLGKETVQNILALRFANQLPDPINW 231

203 ---DDKOKKLTLLTNHRCITRLALTGVEKKIGAAE--VRN-----LLATRPSSQL 249

232 AHYVDHVQ-----ITMAEDIGLGRAGYDYGIGAARDVIONHLMOLLALTAMEEPVSFH 285

250 DSYIMA-KVYVILASGAIIGNP-----QILYNSGFGS----- 278

286 PAALQAEKIKVLSATRLAEPLDQTSRGQYAGWGQGEKVVGLLDEEGFAEDSTTETFAA 345

279 --LQVTPRNDLSIP-----NLGRYITEQPMAPCOIVLRQEFV--DSVRDDPYGLPMWK 327

346 ITLEVDTERRWAGVPFYLRTGKLGRRVTEIALVF-----RRAPHLFPDATMTDELG----- 396

328 EAVAQHIANKPTDALPIPRDPEPQVTTPTTEEHFWHTQIHRDA---FSYGAV-----GPE 380

337 -----TNAMVIRVQ--PDEGVTLRFGSKVPGTAMEVRDVMNDFSYGSAFAEDSPE 444

381 VDSRVIVDL 389

445 AYERLILDV 453

Search completed: October 28, 2003, 16:00:38

Job time : 48 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2003, 15:47:26 ; Search time 26 Seconds

(without alignments)
982.135 Million cell updates/sec

Title: US-09-856-327-2_COPY_76_618

Sequence: 1 NABEGTVPVPGYHKNQET.....IINTLKGTGDKNTGHEHRL 543

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1142.5	39.4	623	1	PROD CORVE
2	1119	38.6	622	1	PROD TRMTU
3	1116	4.0	514	1	G6PD_MICTU
4	104.5	3.6	1337	1	PR2_DROME
5	101	3.5	1010	1	SCA4_RICPA
6	100	3.4	1142	1	ENAM_PIG
7	98	3.4	925	1	W70T_HUMAN
8	97.5	3.4	328	1	LPRX_YERPE
9	97.5	3.4	639	1	AMTG_ASPAK
10	97.5	3.4	1424	1	NCO3_HUMAN
11	97.5	3.4	1536	1	SIN3_YEAST
12	97	3.3	612	1	AMYG_ASPOR
13	96.5	3.3	291	1	MOPR_ARATH
14	96	3.3	921	1	US02_HSV11
15	96	3.3	964	1	YIN0_YEAST
16	95.5	3.3	519	1	AMTH_SACFI
17	95.5	3.3	708	1	HELS_SUTSO
18	95.5	3.3	1257	1	CCAA_BACTU
19	95	3.3	500	1	YDAK_YEAST
20	94.5	3.3	773	1	CDH_FHACH
21	94.5	3.3	855	1	GAP1_SCHPO
22	94.5	3.3	1054	1	R131_ARATH
23	93.5	3.2	406	1	YNOS_YEAST
24	93	3.2	470	1	LEUZ_AZOVI
25	93	3.2	737	1	AMY1_ADAE
26	93	3.2	890	1	GLND_ECOLI
27	93	3.2	890	1	GLND_SALTY
28	93	3.2	955	1	VP2_BTV17
29	93	3.2	1012	1	UBA1_SCHPO
30	92	3.2	376	1	DJBC_MOUSE
31	92	3.2	704	1	RPA3_BOVIN
32	92	3.2	813	1	KRAF_CAEEL
33	92	3.2	890	1	GLND_SALTY

34	92	3.2	1584	1	U104_CAEEL	P23678 caenorhabdi
35	92	3.2	1638	1	DPO3_LACUA	O9cdt7 lactococcus
36	91.5	3.2	425	1	IUCD_ECOLI	P13295 escherichia
37	91.5	3.2	840	1	MGMB_HUMAN	O9u1a3 homo sapien
38	91	3.1	993	1	YAJ1_SCHPO	O09901 schizosach
39	90.5	3.1	847	1	VAV3_HUMAN	O09kwa homo sapien
40	90.5	3.1	1693	1	SAS_DROME	O04164 drosophila
41	90	3.1	463	1	GUNB_BACSP	P29019 bacillus sp
42	90	3.1	487	1	ENGA_CHLPN	O92762 chlamydia p
43	90	3.1	492	1	FTSA_HELPY	O25629 helicobacte
44	90	3.1	606	1	PRIM_MYXKA	P50070 myxococcus
45	90	3.1	737	1	PL02_MOUSE	O9r0b9 mus musculu

ALIGNMENTS

RESULT 1	PROD CORVE	STANDARD	PRT	623 AA
AC	P79076			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Pyranose oxidase precursor (EC 1.1.3.10) (PROD) (Glucose 2-oxidase).			
OS	Coriolus versicolor.			
OC	Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;			
OC	Apophalloporales; Coriolus.			
OX	NCBI_TaxID=57466;			
RN	[1]			
RP	SEQUENCE FROM N.A., SEQUENCE OF 38-63; 92-98; 186-207; 215-225;			
RP	378-411; 491-513 AND 611-621, CHARACTERIZATION, AND SUBUNIT.			
RC	STRAIN=Pe4a; TISSUE=Mycelium.			
RX	MEDLINE=97177816; PubMed=9025322;			
RA	Nishimura I., Okada K., Koyama Y.,			
RT	"Cloning and expression of pyranose oxidase cDNA from Coriolus			
RT	versicolor in Escherichia coli."			
RU	J. Biotechnol. 52:11-20(1996).			
CC	-1- FUNCTION: May play a role in lignin degradation of wood rot fungi			
CC	by supplying H(2)O(2) for the lignin peroxidases. Acts also on D-			
CC	xylose, L-sorbose, D-galactose and 1,5-anhydroglucitol.			
CC	-1- CATALYTIC ACTIVITY: D-glucose + O(2) = 2-dehydro-D-glucose +			
CC	H(2)O(2).			
CC	-1- COFACTOR: FAD; covalent.			
CC	-1- SUBUNIT: Homotrimer.			
CC	-1- SUBCELLULAR LOCATION: Hyphal periplasmic space (Potential).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL: D73369; BAA1119.1; -			
DR	InterPro: IPRO0205; NAD_binding.			
KW	Oxidoreductase; Flavoprotein; FAD; Periplasmic.			
FT	PROPEP 38 623			
FT	CHAIN 38 623 PYRANOSE OXIDASE.			
SQ	SEQUENCE 623 AA; 69495 MW; 5D3FC81B35FA5854 CRC64;			
Query Match	39.4%; Score 1142.5; DB 1; Length 623;			
Best Local Similarity	45.1%; Pred. No. 1e-80;			
Matches 251; Conservative 69; Mismatches 163; Indels 73; Gaps 15;				
OY	13 GYHKKEIFKQIDRFVNVIGALQGVSPVRNQNVPFLDGAWSAPRGSASISGNKP 72			
DB	87 GAHKNTVYQKNIDFVNVIGQGLMSVSPVNTLVIDLPSTSWQA--SSFFVRGSP 144			
OY	73 HOREPENTIAEAATVTRGVGMSTHMTCTPRIRHPMESLPGIGRPKLSPNDPAE-DDKENNE 131			
DB	145 EDDPRLNISGCAVTRVVGMSTHMTCATPRPRDQ-----RPLLVDDQDADDAEMDR 197			

QY 132 LYSEARLIGTSTKEFDESIRHTLVLRSLQDAYKDRQIRFPLPLACHRLKNAPEYVEWH 191
 DB 198 LYTKAESYFKTGTDQFQKESIRHNLVNLKLAEEYKG-QRDFQOIPLAATR--RSPTFVWS 254
 QY 192 SAENLP--HSYNDKOKKLLFTLLTNHRCRLALTGGYEKKIGAAEVRNLLATRNPSQL 249
 DB 255 SANTVFDLQNRPNTPDAPNERENLFPVACERV-----VRN-----TSNSEI 295
 QY 250 DSY-----IMAKVVLASGAINQOILVNSGSL-QVTPRN-DSLIPNLGRY 295
 DB 296 ESLIHDLISGRFEIKADVFLVLTAGAVHNAQLVNSGFGQLGRPDANPPQLPSLSGY 355
 QY 296 ITEQPMAPCOILVLRQEFVDSVRDDPY--GLP-----WKEAQAQH 333
 DB 356 ITEQSLVFCQVMSDELIDSVKSDMIIRGNPDGLGYSVYTPGAETNKHDPWNEKVKNH 415
 QY 334 IAKNPTDALPIPRDPEPOVTTPTTEHPWHTQIHRDAFSYGAQVPEVDSRVIVDLRWF 393
 DB 416 MNOHQEDPLPIPFEDPEPQVTTLFQSPHPWHTQIHRDAFSYGAQVQOQSDSLIVDRFFG 475
 QY 394 ATDPEANLLVQNDVQGYSMPOPTFRYR--PSTASNVRAKMMADMCEVASNLGGYLPT 452
 DB 476 RTEPKENKLFPSDKITDYNMPOPTDFRPPAGRTSKAEADMTDMCWSAKIGGFLPG 535
 QY 453 SPPQFMDPLALHLAGTTRIGFKA--TTVADNNSLVMDFANLYVAGNGTIRTGFGENPT 510
 DB 536 SLPOFMEGLVHLGGTHRMGFDQEDKCCVNTDSRVFGKFLFGCGGNIPATYAGNPT 595
 QY 511 LTSCHAIKARSIN 526
 DB 596 LTAMSLAIKSEYIKN 611

RESULT 2

PROD_TRAHI ID PROD_TRAHI STANDARD; PRT; 622 AA.
 AC P59097;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Pyranose oxidase precursor (EC 1.1.3.10) (PROD) (Glucose 2-oxidase).
 OS Trameetes hirsuta (White-rot fungus) (Coriolus hirsutus).
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 OC Aphyllophorales; Trameetes.
 OX NCBI_TaxID=5327;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-62; 405-412; 486-491 AND 575-596.
 RC STRAIN=DSM 12119;
 RA Christensen S., Lassen S.F., Schneider P.;
 RT "Nucleic acids encoding polypeptides having pyranose oxidase activity.";
 RL Patent number US6146865, 05-MAY-1999.
 CC -!- FUNCTION: May play a role in lignin degradation of wood rot fungi by supplying H(2)O(2) for the lignin peroxidases. Acts also on D-xylose, L-sorbose, D-mannose, D-galactose, 2-deoxy-D-glucose, methyl-alpha-D-glucopyranoside, methyl-beta-D-glucopyranoside, 5-keto-D-gluconic acid, and D-gluconic acid.
 CC -!- CATALYTIC ACTIVITY: D-glucose + O(2) = 2-dehydro-D-glucose + H(2)O(2).
 CC -!- COFACTOR: FAD; covalent (By similarity).
 CC -!- SUBCELLULAR LOCATION: Hypothal periplasmic space (Potential).
 DR InterPro: IPR000205; NAD binding
 KW Oxidoreductase; Flavoprotein; FAD; Periplasmic.
 FT PROPEP 1 37
 FT CHAIN 38 622 PYRANOSE OXIDASE.
 FT CONFLICT 60 60 T -> S (IN REF. 1; AA SEQUENCE).
 SQ SEQUENCE 622 AA; 69082 MW; 0B616C02014E7C5B CRC64;

Query Match 38.6%; Score 1119; DB 1; Length 622;
 Best Local Similarity 44.6%; Pred. NO. 6.8e-79;
 Matches 251; Conservative 69; Mismatches 183; Indels 60; Gaps 15;

QY 13 GYHKKNEIEQKIDRFVNVVVIKALQOVSVVPRNQNVTPLDPAWSAPPOSSAISNGNP 72
 DB 87 GSHKNTVEYQKIDKFNVIQQLMEVSVVPMVMTVDTLSPASWQA--STFFVRNGANP 144
 QY 73 HQREFENLSAEATRGVGGMSTHTCTSPRIHPHMESLPGIGRPKL-SNDPAEDDKWNE 131
 DB 145 EODPLRNLSCQAVTRVVGGMSTHTCATPFEK-----LQRPLLVKNDSKADDAEADR 197
 QY 132 LYSEARLIGTSTKEFDESIRHTLVLRSLQDAYKDRQIRFPLPLACHRLKNAPEYVEWH 191
 DB 198 LYTKAESYFKTGTTQPAESIRHNLVNLKLAEEYKG-VRDFQOIPLAATR--QSPTFVWS 254
 QY 192 SAENLP--HSYNDKOKKLLFTLLTNHRCRLALTGGYEKKIGAAEVRNLLATRNPSQL 249
 DB 255 SANTVFDLQNRPNTPDAPNERENLFPVACERV-----VRN-----TSNSEI 295
 QY 250 DSY-----IMAKVVLASGAINQOILVNSGSL-QVTPRN-DSLIPNLGRY 295
 DB 296 ESLIHDLISGRFEIKADVFLVLTAGAVHNAQLVNSGFGQLGRPDANPPQLPSLSGY 355
 QY 296 ITEQPMAPCOILVLRQEFVDSVRDDPY--GLP-----WKEAQAQH 333
 DB 356 ITEQSLVFCQVMSDELIDSVKSDMIIRGNPDGLGYSVYTPGAETNKHDPWNEKVKNH 415
 QY 334 IAKNPTDALPIPRDPEPOVTTPTTEHPWHTQIHRDAFSYGAQVPEVDSRVIVDLRWF 393
 DB 416 MNOHQEDPLPIPFEDPEPQVTTLFQSPHPWHTQIHRDAFSYGAQVQOQSDSLIVDRFFG 475
 QY 394 ATDPEANLLVQNDVQGYSMPOPTFRYR--PSTASNVRAKMMADMCEVASNLGGYLPT 452
 DB 476 RTEPKENKLFPSDKITDYNMPOPTDFRPPAGRTSKAEADMTDMCWSAKIGGFLPG 535
 QY 453 SPPQFMDPLALHLAGTTRIGFKA--TTVADNNSLVMDFANLYVAGNGTIRTGFGENPT 510
 DB 536 SLPOFMEGLVHLGGTHRMGFDQEDKCCVNTDSRVFGKFLFGCGGNIPATYAGNPT 595
 QY 511 LTSCHAIKARSIN 526
 DB 596 LTAMSLAIKSEYIKN 611

RESULT 3

G6PD MYCTU ID G6PD MYCTU STANDARD; PRT; 514 AA.
 AC O08407;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) (G6PD).
 GN ZWF OR ZWF2 OR RV1447C OR MT1494 OR MTCV493.07.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RC SEQUENCE FROM N.A.
 RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.";
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;

Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayam L.A., Emolaeva M.D., Salzberg S.L., Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula S., Bishai W.;
 "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains";
 Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 -!- CATALYTIC ACTIVITY: D-glucose 6-phosphate + NADP(+) = D-glucose-1,5-lactone 6-phosphate + NADPH.
 -!- PATHWAY: Pentose phosphate pathway; first step.
 -!- SIMILARITY: BELONGS TO THE GLUCOSE-6-PHOSPHATE DEHYDROGENASE FAMILY.
 -!- CAUTION: M.TUBERCULOSIS HAS TWO GENES FOR ZWF. THIS ONE LOOKS LIKE A CLASSICAL ZWF.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 EMBL: Z95844; CAB09259.1; -
 EMBL: AE007019; BAK45757.1; -
 PIR: B70917; B70917.
 HSSP: P11411; IDPG.
 TIGR: MT1494; -
 TrEMBL: Rv1447C; -
 InterPro: IPR001282; G6PD.
 Pfam: PF00479; G6PD.1.
 Pfam: PF02781; G6PD.C.1.
 PRINTS: PR00079; G6PDHGNASE.
 ProDom: PD001129; G6PD; 1.
 TIGRFAMs: TIGR00871; zwf; 1.
 PROSITE: PS00069; G6P DEHYDROGENASE; 1.
 Oxidoreductase; NADP; Glucose metabolism; Complete proteome.
 ACT_SITE 210 210 BY SIMILARITY.
 SEQUENCE 514 AA; 57343 MW; 22A9CCDC9AB062F CRC64;

 Query Match 4.0%; Score 116; DB 1; Length 514;
 Best Local Similarity 22.8%; Pred. No. 0.26;
 Matches 98; Conservative 48; Mismatches 163; Indels 120; Gaps 23;
 29 FVNVKALQV-VSVVNVQNVPTLDGAWAPPGSAISGNKPHOREFENLGAEAATVTR 87
 77 FGQVYNVAQEHCTPPFQQWDLAEG-FRVPCTFDDDAFAQLAETLEKLDAAE---R 132
 88 GVGMSTHTWCTSPRIHPM-ESL--PGIGRPKLSNDPAEDDKENNELYSEAERLIGTST 144
 133 GTGNGHAFYLAIPKSPFVQCEQLHKSGLARPQ-----GDRWSRVVIE----- 175
 145 KEPDESIRHTLVLSLDAYKDRQIRPLPLACHRL--KNAPEVVEVHSAENLPHSTYN 202
 176 KPGHDLASARELNKAVNAVPEEVAEVR-----IDHYLGKTVQNLARFANFQDFPIWN 231
 203 ----DDQKKLFTLLNHRCTRLALTGTGVEKIGAAE--VKN-----LLATNPSSOL 249
 232 AHVVDHVQ-----ITWAEIDIGLGRAGYDGIQARVDVQNLMLQALLTAMEEVSFH 285
 250 DSYTMA-KVYVLSAGATGNP-----QILYNSGFSG----- 278
 286 PAALQAEKIKVLSATRLAELPDLQTTSRQYAAQGGKGVVGLLDEGFAEDSTTETFAA 345
 279 --LQVTRNLSLIP-----NLGRYTEQPMACQIVLREFV--DSVRDDPVLGLPWK 327
 346 ITLEVDFRWRAGVFFYLRTGKRLGRRVTEIALVP-----RRAPHLFPDATMTDELG- 396
 328 EAVAQIAKNTDALPIPFDRPEQVPTPTFTEHPHWTQIHRDA---FSYGAIV---GPE 380
 397 -----TNAMVIRVQ-PDEGVTLRFGSKVPGTAMEVRDVMNDFSYGSAFAEDSPE 444

QY 381 VDSRVIVDL 389
 DB 445 AYERLILDV 453
 RESULT 4
 PR2_DROME STANDARD; PRT; 1337 AA.
 AC Q917F7; Q24316; Q96026; Q9V6K0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Tyrosine-protein kinase PR2 (EC 2.7.1.112).
 GN PR2 OR CG3969.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID:7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkelley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Buesam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mout R.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 [2]
 RN REVISIONS, AND ALTERNATIVE SPLICING.
 RC STRAIN=Berkelley.
 RX MEDLINE=22426063; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celisner S.E. de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a

systematic review.";
Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
[3]
SEQUENCE FROM N.A. (ISOFORM LONG).
RC STRAIN=Berkeley; TISSUE=Embryo;
RA Stapleton M., Brokstein P., Hong L., Agbavani A., Carlson J.W.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Friese E., George R.A.,
RA Gonzalez M., Guarin H., Li P.W., Liao G., Miranda A., Mungall C.J.,
RA Nuno J., Pacleb J.M., Paragas V., Park S., Phouanavong S.,
RA Wan K.H., Yu C., Lewis S.E., Rubin G.M., Celiniker S.E.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
[4]
SEQUENCE OF 13-917 FROM N.A. (ISOFORM LONG).
RC STRAIN=Oregon-R; TISSUE=Embryo;
RX MEDLINE=94156202; PubMed=811807;
RA Ito M., Mateui T., Taniguchi T., Chihara K.;
RT "Alternative splicing generates two distinct transcripts for the
RT Drosophila melanogaster fibroblast growth factor receptor homolog.";
Gene 139:215-218(1994).
[5]
SEQUENCE OF 266-321 FROM N.A.
RX MEDLINE=98401146; PubMed=9731193;
RA Oates A.C., Wollberg P., Achen M.G., Wilks A.F.;
RT "Sampling the genomic pool of protein tyrosine kinase genes using the
RT polymerase chain reaction with genomic DNA.";
RL Biochem. Biophys. Res. Commun. 249:660-667(1998).
CC -|- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -|- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q917F7-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q917F7-2; Sequence=VSP_005006;
CC -|- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES.
CC -|- SIMILARITY: Contains 1 CRIB domain.
CC -|- CAUTION: Ref. 4 sequence differs from that shown due to frameshifts
CC in positions 752, 806, 816 and 827.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE003819; AAF58423.4; -;
CC EMBL; AE003819; AAC22275.3; -;
CC EMBL; AY051755; AAX93179.1; -;
CC EMBL; D17551; BAA04489.1; ALT_FRAME.
CC EMBL; AJ002909; CAA05744.1; -;
CC HSSP; P11362; IFGK.
CC FlyBase; FBgn0013955; P2.
CC GO; GO:0004713; F:protein tyrosine kinase activity; NAS.
CC GO; GO:0006468; P:protein amino acid phosphorylation; NAS.
CC InterPro; IPR000095; FAKbox/KhoBind.
CC InterPro; IPR000719; Prot_kinase.
CC InterPro; IPR001245; Tyr_kinase.
CC Pfam; PF00018; SH3; 1.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Prot_kinase; 1.
CC SMART; SM00285; PBD; 1.
CC SMART; SM00220; S_TKC; 1.
CC SMART; SM00219; TyKc; 1.
CC PROSITE; PS00108; CRIB; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Transferase; Kinase; Tyrosine-protein kinase; ATP-binding;
KW Alternative splicing.
FT DOMAIN 133 399 PROTEIN_KINASE.
FT NP_BIND 139 147 ATP (BY SIMILARITY).

FT BINDING 164 164 ATP (BY SIMILARITY).
FT ACT_SITE 260 260 BY SIMILARITY.
FT DOMAIN 488 502 CRIB.
FT VARSPLIC 1 82 Missing (in isoform Short).
FT CONFLICT 450 450 /FTid=VSP_005006.
FT CONFLICT 484 484 F -> FSTFWKGLSTGKTGYF (IN REF. 4).
FT CONFLICT 906 906 R -> A (IN REF. 4).
FT CONFLICT 913 913 A -> P (IN REF. 4).
FT CONFLICT 913 913 D -> V (IN REF. 4).
SQ SEQUENCE 1337 AA; 147475 MW; 12650E9DF5A82724 CRC64;
Query Match 3.6%; Score 104.5; DB 1; Length 1337;
Best Local Similarity 20.7%; Pred. No. 8.5;
Matches 132; Conservative 77; Mismatches 228; Indels 201; Gaps 34;
QY 5 GRAVPVPGYHKQNEIEFQKIDRFVNVKALQOVSVVRNQNVP-----TLDPGAMS 58
Db 48 GLSRPEIRLRKRYEKHFPHS---YLSKIKRLQLQAPCTVMVKREAPGGGQVALD-GSSA 103
QY 59 APGSSAISNG-----KNPHQREFENLSAEAVTRGVG---GMSH--WTCSTPRIHPM 107
Db 104 SACSSLAANKGASSPSKVPNNKHIIIPADSLSNVKNLQGTGEGFIVQGVWNGNERIOVAI 163
QY 108 ESLPGIGRKLNDPAEDDKENNELYS-EAERLI-----GTSTKEFDESIRHTLVLR 159
Db 164 KCL---CRERMOSNPFLEKAAIMHSIENIVRLYGUVLATDSLMLVTELAH---LRS 217
QY 160 LQDAYKDR--QRIFRPLPLACH---RLKNAPEYVHSAENLFHS-----IYNDKQ 206
Db 218 LLECLKDSGLRVSLFTIPLCEFALQICGMRYLE---QKRLIHRDLAARNILVFSKDKV 274
QY 207 KKLFTLLTHRCTR-LALTGGYEK-----KIG-----AAEVNLLATRNPSQLDSYIM 254
Db 275 K-----ISDFGLSRALGVGKDYKTNPNVNLKLIPIAWCAPECINYLRFTHASD----- 322
QY 255 AKVYVLAASGAIGNPQILYN---SGFSGLOVTPRNDLSLIPNLGRYITEQP----- 300
Db 323 ----VWAFGVCLWEMPSYGFQWPAALTGLQILEAIDA--PNYOR--LEQDDCCPSEYTL 374
QY 301 MAPC-----QIVLRQEFVDSVRDDPYG 322
Db 375 MNKQWDDAAKPRFGEIYDQPMKPEQLKAVVNVCTEPKKDHLRYQGDIIISVLDRTNG 434
QY 323 LPWKEAVAQHAQAK--NPT-----DALPIPRDPPEOVTPPTTEHPHWTOI-----HRD 370
Db 435 TPFWKGLVLTGTGYVNPNTVAFLGLPSSTRDSFVSVDHRSRKRKATEMISKPDND 494
QY 371 AFSYGAVGPEVDSRVIVDLRWFGAT-----DPEANNLLVFQNDVQDG 412
Db 495 FKHTGHVG--IDGATFGDIAFLGSSQNYNHVPKQIVTPKPSEDIQOTPLL----- 544
QY 413 YSMPOPTFRYRPSTASNVNRKMMADMCEVASNLGGYLPSTPPQFMDPGLALHLAGTTRI 472
Db 545 --PPTPTSPDSLQTSAGYFPE-----GANSNGAMGTS-----KNPTFIPSAEHTPKL 589
QY 473 GFDKATTVADNNSLVWDFA-----NLTVAGNGTIRTCFG 506
Db 590 -----IATNGQSSPFDFAAGSTNPFNPRGDDLEFG 620
RESULT 5
SCA4_RICPA STANDARD; PRT; 1010 AA.
ID SCA4_RICPA
AC Q9AJ75;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
DE (Protein PS 120) (fragment).
GN SCA4 OR D.
OS Rickettsia parkeri.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiaseae; Rickettsia.

FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 264 264 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 291 291 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CARBOHYD 462 462 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CARBOHYD 929 929 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CARBOHYD 1040 1040 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CONFLICT 680 680 H -> D (IN REF. 2)
 FT CONFLICT 838 840 RDH -> TTI (IN REF. 2)
 SQ SEQUENCE 1142 AA; 128352 MW; 938306BC87CC3FC6 CRC64;

Query Match 3.4%; Score 100; DB 1; Length 1142;
 Best Local Similarity 19.2%; Pred. No. 15;
 Matches 118; Conservative 65; Mismatches 210; Indels 220; Gaps 32;

QY 1 NAEGETAVPYV-GYH- KNEIEPKQIDR- FVWIKG 35
 DB 192 NEEGNN-PYRGFGYHGGGRPPYSEEMFEQDFEKPKEKDPPTETPATESVN- 245
 QY 36 ALQOVSVVRNVPYL-DP-GAWSAPPSSAISGNKNPHQR- EFENLSABAVTR 87
 DB 246 -TTVPETNSTQPNAPNPGNDTSPTGTS--CGGNPRSNPTGONGPAVNVSGQGVPR 299
 QY 88 -GGVGMSTHWTCTSPRI- HPPMESLFGIGR- PKLSNDPAEDDKW 129
 DB 300 SQSPWGPRTQIHENYENPNIRGFPARQWRPPG--PAMGHRNGFPYRNQIQORGRW 356
 QY 130 NELYSEABRLIGTSKTFDESIRHTLVLSLODAYKORIFRPLPLACHRLKNAPYVE 189
 DB 357 NSFTLEGG- QAVRPGYPTIRYVY- GSTARSNPPNYA- 390
 QY 190 WSAENLFHSIYNDKQKLPFLTLNHRCTRLALTG- GYEKKIGAAE 235
 DB 391 -GNSANLRRKEGPNK- NPMVTNVAIPGPKHGTVDQENIQNPREKQVSK 440
 QY 236 VRNLLATRNPS- SOLDSYIMAKVYVLASGAINPQILYNSGFGSLQVTPRNDLSIPN 291
 DB 441 -RTVVPTRDPGWRNSODYGINSKYKL- PQ- PEDNMLVPN 479
 QY 292 LG- RYTEOPMAFCQIVLQEFVDSVRDDPYGLPWKEA- VAQ 332
 DB 480 FNSIDORENSYYPGESKRAPNSDGTQTQIIPK- GIVLEPRPIPYESETNOPELK 534
 QY 333 HIAKNP-TDALPIPRDPPEP- QVTPTEEHPWHQIHRDAFSGVAVGPEV 381
 DB 535 HSAVQPVVTEGIPSPAKEHPPAGNTWNQOEISPPFKED-PGQEEHLPHLSHG- 587
 QY 382 DSRVIVDLRFMGATDPEANLLVFQNDVQDYSMP- OPTFYRPPSTASNVKRWMA 437
 DB 588 -SRVHVYPDNPYDPRENSPYLSRNTWYERDDSPNTMGQPENPHYPMTPDPKET- 642
 QY 438 DMCEVASNLGYLTPSPQPMQDGLALHLACTTTRIGFDKATTVADNNSLVWDFANLYVAG 497
 DB 643 -IPYNEEDPIDPTGDEHFPQQR- WDMBEELSFK 674
 QY 498 NGTIRTCGGENPT 510
 DB 675 DPTVRHYEGEQT 687

RESULT 7

W70T_HUMAN
 ID W70T_HUMAN STANDARD; PRT; 925 AA.
 AC P57737;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 70 kDa WD-repeat tumor rejection antigen homolog.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
 Tanaka T., Nakamura Y., Isogai T., Sugano S.;
 "NEDO human cDNA sequencing project";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: Contains 8 WD repeats.
 CC -!- SIMILARITY: BELONGS TO THE CORONIN FAMILY OF WD-REPEAT PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AK025674; BAB15211.1; -;
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 6.
 DR PRINTS; PR00320; GPROTEINRPT.
 DR SMART; SM00320; WD40; 7.
 DR PROSITE; PS00678; WD_REPEATS_1; 1.
 DR PROSITE; PS50082; WD_REPEATS_2; 4.
 DR PROSITE; PS50294; WD_REPEATS_REGION; 2.
 KW Repeat; WD repeat.
 FT REPEAT 75 115 WD 1.
 FT REPEAT 124 163 WD 2.
 FT REPEAT 166 205 WD 3.
 FT REPEAT 209 253 WD 4.
 FT REPEAT 542 582 WD 5.
 FT REPEAT 592 632 WD 6.
 FT REPEAT 635 674 WD 7.
 FT REPEAT 728 768 WD 8.
 SQ SEQUENCE 925 AA; 100574 MW; A393CE973C94F4A9 CRC64;

Query Match 3.4%; Score 98; DB 1; Length 925;
 Best Local Similarity 20.2%; Pred. No. 16;
 Matches 88; Conservative 52; Mismatches 139; Indels 156; Gaps 21;

QY 54 PGAWSAPPSSAISGNKNPHQREFENLSABAVTRGV- GGMSTHWTCTPRIHPME 108
 DB 108 PGQALPSAGPVVLG- PEDLPVEVLFQFHTSDGILVSAAGTTVKWDAKQ- PLT 162
 QY 109 SLPGICRPKLSNDPAEDDKENNELYSEAEELICTSTKEFDESIRHTLVLSLODAYKDRQ 168
 DB 163 EL- AHGDLVQSAVSRDGLVGT- ACKDQ 191
 QY 169 -RIFRPLPLACHRLK- NAPEYVEWH- SAENLFHSIYNDKQK- 209
 DB 192 LRIFDP- RTKPRASQSTOAHENSDRSRLAWGTWEHLVSTGNGREREVKLWD 245
 QY 210 -FTLTNRHCT- RIALTGGYEKKIGAAEVRNLLATRNPSQ- L 249
 DB 246 RPFSSALASITLDTSLGCLVPLDPSGLLVLAGKGRQLCYEVVPPQALSPVTCVL 305
 QY 250 DSYIMAKVYVLASGAINPQILYNSGFGSLQVTPRNDLSIPNLGRYTEQPMACQIVLR 309
 DB 306 ES- VLRGAALVPRQALAVMGCEVLRLQLSDTAIVPIGYHVRKAVEFHEDL- 356
 QY 310 QEFVDSV- RDDPYGLPW- KEAVAQHIKNP- TDALPIPRDPPEQVTT 355
 DB 357 -FPDTAGCVATDPDTHG- WAGDNQOVQKVSLNPACRPHPSFTSCL-VPAEPLPTAQ 411
 QY 356 PFTEEHPWHQIHRDAFSGVAVGPEVDSRVIVDLRFMGATDPEANLLVFQNDVQDYSMP 415
 DB 412 PAVMETP- VGDADASEGFSS 430
 QY 416 PQTFYRPPSTASNV 430
 DB 431 P-PSLTSPTSPSSL 444

RESULT 8

LPXK_YERPE
ID LPXK_YERPE STANDARD; PRT; 328 AA.
AC Q82GAG;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Tetraacyldisaccharide 4'-kinase (EC 2.7.1.130) (Lipid A 4'-kinase).
GN LPXK OR YPO1396 OR Y2776.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586350;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Felwell T., Hamlin N., Holtroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RA "Genome sequence of Yersinia pestis, the causative agent of plague";
RA Nature 413:523-527(2001).
IL [2]
IN SEQUENCE FROM N.A.
IP STRAIN=KIMS / Biovar Mediaevalis;
IC MEDLINE=22137863; PubMed=12142430;
JC Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
JA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
JA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
JA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
JA Perry R.D.;
JA "Genome sequence of Yersinia pestis KIM";
JA J. Bacteriol. 184:4601-4611(2002).
LT J. FUNCTION: Transfers the gamma-phosphate of ATP to the 4'-position
of a tetraacyldisaccharide 1-phosphate intermediate (termed DS-1-
P) to form tetraacyldisaccharide 1,4'-bis-phosphate (lipid IVA)
(By similarity).
CT -1- CATALYTIC ACTIVITY: ATP + 2,3-bis(3-hydroxytetradecanoyl)-D-
glucosaminyl-(beta-D-1,6)-2,3-bis(3-hydroxytetradecanoyl)-D-
glucosaminyl beta-phosphate = ADP + 2,3,2',3'-tetrakis(3-
hydroxytetradecanoyl)-D-glucosaminyl-1,6-beta-D-glucosamine 1,4'-
bisphosphate.
CT -1- PATHWAY: Lipid A biosynthesis; sixth step.
CT -1- SIMILARITY: BELONGS TO THE LPXK FAMILY.
CT This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CT ENBL; AJ414148; CAC90225.1;
CT EMBL; AB013880; AAM86328.1;
CT PIR; AF0170; AF0170.
CT HAMAP; MF_00409; -; 1.
CT InterPro; IPR003758; LPXK.
CT Pfam; PF02606; Lpxk; 1.
CT TIGRFAMs; TIGR00682; lpxk; 1.
CT Transferase; Kinase; Lipid A biosynthesis; Lipid synthesis;
CT ATP-binding; Complete proteome.
CT NP BIND 55 62 ATP (POTENTIAL).
CT SEQUENCE 328 AA; 36138 MW; 46E96728C2624CCA CRC64;
Query Match 3.4%; Score 97.5; DB 1; Length 328;
Best Local Similarity 23.2%; Pred. No. 3.7;
Matches 74; Conservative 131; Indels 75; Gaps 16;
56 AWSAP-----PGSSAISNGKNP-----HQREFENLSAEAVTRGVGGMSTHW----- 96

Db 41 AWRSPVVIIVGNLTAGNGKTFVIMVLEQLQGRVYGVSRGYSKSAVYPLLLSDN 100
Qy 97 TCSPTRIHPMESLPGIGRKLNDPAEDDKENNELYSEARLIGTSTKFEDESIRHTLV 156
Db 101 TTTAAQAGDEPVLIFORTGAP-VAVSPKRAAIKALQSHAVDFIIT-----DGLQHYAL 154
Qy 157 LRSLODAYKDRQRIFRP---LPLACHRLKNAPEYVWHSANLFSHYNDKOKKLFLL 213
Db 155 QRFDFELVVDGVRFRFGVWLPAGPMR-----EGRLSVDAAI 194
Qy 214 TNRCTRLALTGGYEKKIGAAEVRNLL-ATRNSSOLDVYIMAKVYVYLAAGNPOILY 272
Db 195 TN---GGLAAGEIPMLQVAREAVNLVTGQPAEQQLQ-----HVVAMAGIHPRRFF 244
Qy 273 NSGFSGLQVTPRN-----DSLIPNLGRYTEQPMACQIVLROEFVDSVDDPYGLP 324
Db 245 -ATLNLGKIPENEHAFADHQDYSLAQLSR-LTSGP---QILLMTE-KDAVKRAFALP 297
Qy 325 -WVKEAVAQIAKNPTDAL 342
Db 298 NWYLPVDAQLPSSDRADKL 316
RESULT 9
AMYG ASPAK STANDARD; PRT; 639 AA.
ID AMYG ASPAK
AC P23176;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glucoamylase I precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
DE (1,4-alpha-D-glucan glucohydrolase).
GN GAI.
OS Aspergillus awamori (var. kawachi).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OX Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=40384;
RN [1]
RP SEQUENCE FROM N.A.
RA Hayashida S., Kuroda K., Ohta K., Kuhara S., Fukuda K., Sakaki Y.;
RT "Molecular cloning of the glucoamylase I gene of Aspergillus awamori
var. kawachi for localization of the raw-starch-affinity site";
RL Agric. Biol. Chem. 53:923-929(1989).
RN [2]
RP SEQUENCE OF 494-538
RA Hayashida S., Nakahara K., Kuroda K., Miyata T., Iwanaga S.;
RT "Structure of the raw-starch-affinity site on the Aspergillus awamori
var. kawachi glucoamylase I molecule";
RL Agric. Biol. Chem. 53:135-141(1989).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
glucose residues successively from non-reducing ends of the chains
with release of beta-D-glucose.
CC -1- SIMILARITY: BELONGS TO FAMILY 15 OF GLYCOSYL HYDROLASES.
CT This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CT ENBL; D00427; BAA00331.1;
CT PIR; J0479; J0479.
CT PDB; 1GAI; 17-AUG-96.
CT PDB; 1GLM; 31-JUL-94.
CT PDB; 3GLY; 01-NOV-94.
CT GlycoSuiteDB; P23176;
CT InterPro; IPR02044; CBD_4.
CT Pfam; PF00686; CBM_20; 1.
CT Pfam; PF00723; Glyco_hydro_15; 1.

DR PRINTS: PRO0736; GLHYDELASE15.
 DR PRODOM: PD001568; CBD_4; 1.
 DR PROSITE: PS00820; GLUCOAMYLASE; 1.
 KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
 FT SIGNAL; 3D-structure.
 FT SIGNAL 1 18
 FT PROPEP 19 24
 FT CHAIN 25 639
 FT DOMAIN 494 538
 FT BINDING 143 143
 FT ACT SITE 199 199
 FT ACT_SITE 202 202
 FT ACT_SITE 203 203
 FT ACT_SITE 233 236
 FT DISULFID 245 472
 FT DISULFID 285 293
 FT CARBOHYD 194 194
 FT CARBOHYD 418 418
 FT CARBOHYD 464 464
 FT CARBOHYD 466 466
 FT CARBOHYD 467 467
 FT CARBOHYD 475 475
 FT CARBOHYD 476 476
 FT CARBOHYD 482 482
 FT CARBOHYD 483 483
 FT CARBOHYD 485 485
 FT CARBOHYD 487 487
 FT CARBOHYD 491 491
 FT CARBOHYD 495 495
 FT CARBOHYD 498 498
 FT CARBOHYD 499 499
 FT CARBOHYD 500 500
 FT CARBOHYD 501 501
 FT CARBOHYD 503 503
 FT CARBOHYD 505 505
 FT CARBOHYD 507 507
 FT CARBOHYD 511 511
 FT CARBOHYD 512 512
 FT CARBOHYD 513 513
 FT CARBOHYD 514 514
 FT CARBOHYD 516 516
 FT CARBOHYD 517 517
 FT CARBOHYD 519 519
 FT CARBOHYD 521 521
 FT CARBOHYD 523 523
 FT CARBOHYD 524 524
 FT CARBOHYD 525 525
 FT CARBOHYD 526 526
 FT CARBOHYD 527 527
 FT CARBOHYD 528 528
 FT CARBOHYD 529 529
 FT CARBOHYD 530 530
 FT CARBOHYD 531 531
 FT CARBOHYD 533 533
 FT CARBOHYD 534 534
 FT CARBOHYD 639 AA; 68271 MW; E112B31A4D8D06B CRC84;
 SQ SEQUENCE 3.4%; Score 97.5; DB 1; Length 639;
 Query Match
 Best Local Similarity 20.2%; Pred. No. 9.9;
 Matches 106; Conservative 51; Mismatches 194; Indels 175; Gaps 25;
 QY 80 LSAEAVTRGVGGMSTHWTCTPPIHPMPESLPDGPRLKNSDPAEDDKENNELYSAERL 139
 DB 106 ISSQAIIGVSNPSGDL-----SSGGLGEPKFNVDETAYTGTSGMRQPDGPAL 153
 QY 140 IGTSTKEFDESIRHTLVRLSLQDAYKD-RQRIERPLPLACHRLKNAPEYVE--WHAENL 196
 DB 154 RATAMIGFGQVL-----LDNGYTSAAATEIWNPL-----VRNOLSYVAQWYNTG-- 197
 QY 197 FHSYINDKQKQLFTLLTNHRCRLALTGG--YEKIGIAAEVRNLLATRNPPSSQLDSYIM 254
 DB 198 -YDLMEVNGSSPFTIAVQHR-----ALVEGSATATVAGSS-----CSWCDSQA- 240

QY 255 AKVYVLASGAIGNPOIL--YNSGFSGLQVTPRNDSLIPN-----LGRVITEQPMAPCQ- 305
 DB 241 -----PQLLCVLQFWTGSYILANPDSRRSGKDTNLTGSHTFDFPACDD 287
 QY 306 -----IVLRQEFVDSVR-----DDPY--GLPWKCAVAQ 332
 DB 288 STFQPCSPRALANHKHEVDSFRSIVTLNDGLSDSEAVAVGRYPEDSYNGNPNFQSTLA- 346
 QY 333 HIAKMPDTPALPIPRDPPEQVTPPTTEHPWHTQ-----IHRDAFSYGAVG--- 378
 DB 347 -AAEQLYDAL-----YQWQKQSGLEITDVSDFKALYSGAATGYS 387
 QY 379 --PEVDSRVIVDLRWFG-----ATDPBANLLVFQNDVQDG--XSMPOPTFRYRPSA 427
 DB 388 SSSSYSSIVSAVKTFADGFVSIETHAASNSLSQDFKSDGDELSARDLTSYAALLT 447
 QY 428 SNVRARKMM-----ADMCEVASNLGY-----LPTSPQFMDPGLALHLAGTTRI 472
 DB 448 ANNRNRNVVPPSWGETSASSWPGTCAATSAGTSYSSVTVTWSPSIVATGGTTTTATTGTS 507
 QY 473 G----FDKATTVADNNLSLWDFAN-----LYVAGNGTIRTGGEN 508
 DB 508 GGVTSKTTTTSKTTTTSSTTSCTTPTTAVAVTFLDTATTYGEN 553
 RESULT 10
 NC03 HUMAN
 ID NC03 HUMAN STANDARD; PRT; 1424 AA.
 AC Q9Y6Q9; Q9BR49; Q9UPG9; Q9UPG4; Q9UPG7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Nuclear receptor coactivator 3 (EC 2.3.1.48) (NCoA-3) (Thyroid hormone
 DE receptor activator molecule 1) (TRAM-1) (ACTR) (Receptor-associated
 DE coactivator 3) (RAC-3) (Amplified in breast cancer-1 protein) (AIB-1)
 DE (Steroid receptor coactivator protein 3) (SRC-3) (CBP-interacting
 DE protein) (PCIP)
 GN NC03 OR TRAM1 OR RAC3 OR AIB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND INTERACTION WITH CREBBP;
 RP PCAF; RARA; RXRA; THRA AND ESR.
 RX TISSUE=Pituitary;
 RX MEDLINE=98010595; PubMed=9346901;
 RA Takeehita A., Cardona G.R., Koibuchi N., Suen C.-S., Chin W.W.;
 RT "TRAM-1, a novel 160-kDa thyroid hormone receptor activator molecule,
 RT exhibits distinct properties from steroid receptor coactivator-1.";
 RL J. Biol. Chem. 272:27629-27634(1997).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4), ENZYMATIC ACTIVITY, AND VARIANT
 RP 1248-GLN-GLN-1250 DEL.
 RC TISSUE=Leukemia;
 RX MEDLINE=97410321; PubMed=9257036;
 RA Chen H., Lin X.-J., Schiltz R.L., Chakravarti D., Nash A., Nagy L.,
 RA Privalsky M.L., Nakatani Y., Evans R.M.;
 RT "Nuclear receptor coactivator ACTR is a novel histone
 RT acetyltransferase and forms a multimeric activation complex with p/CAF
 RT and CBP/p300.";
 RL Cell 90:569-580(1997).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 5), AND INTERACTION WITH ESR.
 RC TISSUE=Lung;
 RX MEDLINE=97400625; PubMed=92523239;
 RA Anzick S.L., Kononen J., Walker R.L., Azorsa D.O., Tanner M.M.,
 RA Guan X.-Y., Sauter G., Kallioniemi O.-P., Trent J.M., Meltzer P.S.;
 RT "AIB1, a steroid receptor coactivator amplified in breast and ovarian
 RT cancer.";
 RL Science 277:965-968(1997).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 5), INTERACTION WITH VDR; RARA; PPARA;

RXRA; THRA AND ESR, AND VARIANT 1248-GLN--GLN-1250 DEL.
 TISSUE=Brain;
 MEDLINE=97385128; PubMed=9238002;
 Li H., Gomes P.J., Chen J.D.;
 "RAC3, a steroid/nuclear receptor-associated coactivator that is
 related to SRC-1 and TIF2.";
 Proc. Natl. Acad. Sci. U.S.A. 94:8479-8484(1997).
 [5]
 SEQUENCE FROM N.A.
 MEDLINE=21638749; PubMed=11780052;
 Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,
 Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beate D.M.,
 Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 Chapman J.C., Clamp M., Clark L.N., Clark S.Y., Clee C.M.,
 Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
 Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 Grahame D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 Lehaeaslaio M.H., Leverisha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 Marsh V.L., Martin S.L., McConachie L.J., McIay K., McMurray A.,
 Milne S.A., Mistry D., Moore M.J.P., Mullikin J.C., Nickerson T.,
 Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S.,
 Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
 Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 Rogers J.;
 "The DNA sequence and comparative analysis of human chromosome 20.";
 Nature 414:865-871(2001).
 [6]
 ACETYLATION BY CREBBP, AND MUTAGENESIS OF LYS-616; 619-LYS--LYS-620;
 LYS-647; LYS-681; LYS-687; LYS-700 AND LYS-708.
 MEDLINE=99418638; PubMed=10490106;
 Chen H., Lin R.J., Xie W., Wilpitz D., Evans R.M.;
 "Regulation of hormone-induced histone hyperacetylation and gene
 activation via acetylation of an acetylase.";
 Cell 98:675-686(1999).
 [7]
 INTERACTION WITH NFKB1.
 MEDLINE=20547534; PubMed=11094166;
 Werbach S., Nojek I., Lanz R., Costas M.A.;
 "RAC-3 is a NF-kappa B coactivator.";
 PERS Lett. 485:195-199(2000).
 [8]
 SUBUNIT OF A COMPLEX CONTAINING CREBBP, NCOA2, IKKA, IKKB AND IKKKG,
 AND PHOSPHORYLATION.
 MEDLINE=21968797; PubMed=11971985;
 Wu R.-C., Qin J., Hashimoto Y., Wong J., Xu J., Tsai S.-Y., Tsai M.-J.,
 O'Malley B.W.;
 "Regulation of SRC-3 (pICP/ACTR/AIB-1/RAC-3/TRAM-1) coactivator
 activity by I kappa B kinase.";
 Mol. Cell. Biol. 22:3549-3561(2002).
 [9]
 VARIANTS IN POLY-GLN REGION.
 MEDLINE=98394734; PubMed=9727751;
 Shirazi S.K., Bober M.A., Coetzee G.A.;
 "Polymorphic exonic CAG microsatellites in the gene amplified in
 breast cancer (AIB1 gene)."
 Clin. Genet. 54:102-103(1998).
 -1- FUNCTION: Nuclear receptor coactivator that directly binds nuclear
 receptors and stimulates the transcriptional activities in a
 hormone-dependent fashion. Plays a central role in creating a
 multisubunit coactivator complex, which probably acts via
 remodelling of chromatin. Involved in the coactivation of different
 nuclear receptors, such as for steroids (GR and ER), retinoids

(RARs and RXRs), thyroid hormone (TRs), vitamin D3 (VDR) and
 prostanoids (PPARs). Displays histone acetyltransferase activity.
 Also involved in the coactivation of the NF-kappa-B pathway via
 its interaction with the NFKB1 subunit.
 -1- CATALYTIC ACTIVITY: Acetyl-CoA + histone = CoA + acetyl-histone.
 -1- ENZYME REGULATION: Coactivator activity on nuclear receptors and
 NF-kappa-B pathways is enhanced by various hormones, and the TNF
 cytokine, respectively. TNF stimulation probably enhances
 phosphorylation, which in turn activates coactivator function. In
 contrast, acetylation by CREBBP apparently suppresses coactivation
 of target genes by disrupting its association with nuclear
 receptors.
 -1- SUBUNIT: Present in a complex containing NCOA2, IKKA, IKKB, IKKKG
 and the histone acetyltransferase protein CREBBP. Interacts with
 PCAF.
 -1- SUBCELLULAR LOCATION: Mainly cytoplasmic and weakly nuclear. Upon
 TNF activation and subsequent phosphorylation, it translocates
 from the cytoplasm to the nucleus.
 -1- ALTERNATIVE PRODUCTS:
 Event=Alternative splicing; Named isoforms=5;
 Comment=Additional isoforms seem to exist;
 Name=1;
 IsoId=Q9Y6Q9-1; Sequence=Displayed;
 Name=2;
 IsoId=Q9Y6Q9-2; Sequence=VSP_003407;
 Name=3;
 IsoId=Q9Y6Q9-3; Sequence=VSP_003405, VSP_003407, VSP_003408;
 Name=4;
 IsoId=Q9Y6Q9-4; Sequence=VSP_003405, VSP_003406, VSP_003407,
 VSP_003408;
 Name=5;
 IsoId=Q9Y6Q9-5; Sequence=VSP_003408;
 -1- TISSUE SPECIFICITY: Widely expressed. High expression in heart,
 skeletal muscle, pancreas and placenta. Low expression in brain,
 and very low in lung, liver and kidney.
 -1- DOMAIN: Contains three leu-Xaa-Xaa-leu-Leu (LXXLL) motifs. Motifs
 1 and 2 are essential for the association with nuclear receptors,
 and constitute the RID domain (Receptor-interacting domain).
 -1- PTM: Phosphorylated by IKK complex and acetylated by CREBBP.
 Acetylation occurs in the RID domain, and disrupts the interaction
 with nuclear receptors. Both modifications regulate its function.
 -1- POLYMORPHISM: The length of the poly-Gln region is polymorphic in
 the normal population.
 -1- MISCELLANEOUS: NCOA3 is frequently amplified or overexpressed in
 breast and ovarian cancers.
 -1- SIMILARITY: BELONGS TO THE SRC/P160 FAMILY OF NUCLEAR RECEPTOR
 COACTIVATORS.
 -1- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.
 -1- SIMILARITY: Contains 1 PAS (PER-ARNT-SIM) dimerization domain.
 -1- CAUTION: Ref.5 (CAC36067) sequence differs from that shown due to
 erroneous gene model prediction.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; AF016031; AAC51849.1; -
 EMBL; AF036892; AAB92368.1; -
 EMBL; AF012108; AAC51677.1; -
 EMBL; AF010227; AAC51663.1; -
 EMBL; AL034418; CAC17693.1; -
 EMBL; AL034418; CAC36067.1; ALT_SEQ.
 PIR; T03851; T03851.
 PDB; 1KBH; 06-FEB-02.
 Genew; HGNC:7670; NCOA3.
 MIN; 601937; -
 GO; GO:0005634; C:nucleus; NAS.
 GO; GO:0016923; F:ligand-dependent thyroid hormone receptor i. . .; NAS.
 GO; GO:0003713; F:transcription co-activator activity; NAS.

DR CO; GO:0006350; P:transcription; NAS.
 DR InterPro; IPR001092; HUH_Basic.
 DR InterPro; IPR000014; PAS_domain.
 DR Pfam; PF00010; HUH; 1.
 DR Pfam; PF00989; PAS; 1.
 DR SMART; SM00353; HUH; 1.
 DR SMART; SM00091; PAS; 1.
 DR PROSITE; PS00038; HUH_1; FALSE_NEG.

Query Match 3.4%; Score 97.5; DB 1; Length 1424;
 Best Local Similarity 19.1%; Pred. No. 33;
 Matches 102; Conservative 64; Mismatches 210; Indels 159; Gaps 23;

QY 41 SVVNRNVPLDP-----GMSAPPSSAISKPH-QREFEN-----LSAEA 84
 Db TLPLRSNIPGARPVLOQQOQMLQWRGEIPMGANGPYQAAASNLGSLWPDGMLSMQ 1019
 QY 85 VTRGVGMSHTWCTSTRIHPMESLPGIGPKLSNDPAEDDKENNELYSEAEELIGTST 144
 Db 1020 V-----SHGTQNRPLRNSLDL--VGPPSLGQSDERALLDQLHTLLSNWDATGL 1069
 QY 145 KEFD-----ESIRHTLVRLSDAYKDRQIFR-----PLPLACHRL 181
 Db 1070 EEDRALGIPLVNQGALEPKQAFQGEAAVMMDQKAGLYGQTYPAQGPWGGFHLQ 1129
 QY 182 KNAPEYVEVHSAENLFHSIYNDKQKLFLLTNHRTLRALTCGYEKKIAGAEVRNLLA 241
 Db 1130 GQSPS-----FNSMMNQMGQNFPLQGMH-----PRANIMR 1161
 QY 242 TR-NPSSQLDSYIMAKYVVLASGAIGNPILYNSGSLQVTPNDLSLIPNLGYITEOP 300
 Db 1162 PRTPKQLRMQLQRL-----QCGQFLNGSQAELKEMN-----PTAGGAAVWRP 1208
 QY 301 MAFQCIIVLRQEFVDSVRDDPYGLPWEKEAVAQIAKNTDALPIPR----- 347
 Db 1209 MMQFQVSSQGFN-----AQWVQRSELLSHHFQQRVAMMMQOQQQ 1252
 QY 348 ---DPRFQVTPFTEHPMTQHHRDAPSGVAGPEVDSRVIVDLRWFGATDPEANNLLV 404
 Db 1253 QQQ 1307
 QY 405 FQNDVQGYSM---PQTFRVRPTASVNRKVMADCEVASNLGSLYLTSPQFNDPG 461
 Db 1308 -QFPYQPNYMGQQQDFAFG-RVSSPPNAM-----MSSRMG---FSQNPMMQHPQ 1352
 QY 462 LALHLAGTTRIGFKATTVADNNSLVMDFANLYVAGNGTIRTGFGENTILTSMSCH 516
 Db 1353 AASIYQSSEMKGWFSG-NLARNSS---FSQQQFAHQG-----NPAYYSVWH 1394

RESULT 11
 SIN3 YEAST
 ID SIN3 YEAST STANDARD; PRT; 1536 AA.
 AC P22579; Q08049;
 DT 01-AUG-1991 (rel. 19, Created)
 DT 30-MAY-2000 (rel. 35, Last sequence update)
 DT 30-MAY-2000 (rel. 39, Last annotation update)
 DE Paired amphipathic helix protein.
 GN SIN3 OR SD11 OR UME4 OR RPD1 OR GAM2 OR SDS16 OR YOL004W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GRF88.
 RA MEDLINE=91042523; PubMed=2233725;
 RT Wang H., Clark I., Nicholson P.R., Herskowitz I., Stillman D.J.;
 RT "The Saccharomyces cerevisiae SIN3 gene, a negative regulator of HO,
 RL contains four paired amphipathic helix motifs.";
 RN Mol. Cell. Biol. 10:5927-5936 (1990).
 RP SEQUENCE FROM N.A.

RA Hughes B., Pohl T.M.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR BOTH FULL TRANSCRIPTION REPRESSION AND
 CC ACTIVATION OF MANY GENES INCLUDING CELL TYPE-SPECIFIC GENES (STE6,
 CC TY2 AND HO), CELL DIFFERENTIATION-SPECIFIC GENES (SPO13), GENES
 CC THAT RESPOND TO EXTERNAL SIGNALS (PHO5) AND TRK2. IT IS PROBABLY
 CC INVOLVED IN THE SAME TRANSCRIPTIONAL REGULATORY FUNCTION OR
 CC PATHWAY AS THE TRANSCRIPTIONAL REGULATORY PROTEIN RPD3
 CC -1- SUBUNIT: IT PROBABLY FORMS A COMPLEX WITH THE TRANSCRIPTIONAL
 CC REGULATORY PROTEIN RPD3.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR, POSSIBLY LINKED TO CENTROMERE.
 CC -1- DOMAIN: CONTAINS 4 PAIRED AMPHIPATHIC HELIX MOTIFS, SEPARATED BY A
 CC 10 TO 30 AA SEGMENT THAT FORMS POSSIBLY A LOOP, RESULTING IN A
 CC STRUCTURE SIMILAR TO THAT OF HLH AND TPR MOTIFS.
 CC -1- SIMILARITY: TO S-POMBE SPAC12C2.10C.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).

DR EMBL; M36822; AAA34839.1; --
 DR EMBL; Z74746; CAA99003.1; --
 DR PIR; S66686; RGSYS3.
 DR TRANSFAC; T03597; --
 DR SGD; S0005364; SIN3.
 DR GO; GO:000118; C:histone deacetylase complex; IPI.
 DR GO; GO:0004407; F:histone deacetylase activity; IPI.
 DR GO; GO:0006347; P:chromatin silencing at RML and HMR (sensu S.); IMP.
 DR GO; GO:000183; P:chromatin silencing at ribosomal DNA (rDNA); IMP.
 DR GO; GO:0006348; P:chromatin silencing at telomere; IMP.
 DR GO; GO:0006357; P:regulation of transcription from Pol II pro.; IGI.
 DR InterPro; IPR003822; PAH.
 DR Pfam; PF02671; PAH; 3.
 KW Repeat; Transcription regulation; Repressor; Cell division;
 KW Activator; Nuclear protein.
 FT DOMAIN 239 1200 4 X 2 PAIRED (A,B) AMPHIPATHIC HELICES.
 FT REPEAT 239 252 HELIX A.
 FT REPEAT 274 285 HELIX B.
 FT REPEAT 426 439 HELIX A.
 FT REPEAT 461 472 HELIX B.
 FT REPEAT 679 692 HELIX A.
 FT REPEAT 714 725 HELIX B.
 FT REPEAT 1152 1165 HELIX A.
 FT REPEAT 1189 1200 HELIX B.
 FT DOMAIN 480 519 GLN-RICH.
 FT CONFLICT 510 510 Q -> QAQ (IN REF. 2).
 SQ SEQUENCE 1536 AA; 174838 MW; 0834726312B13878 CRC64;

Query Match 3.4%; Score 97.5; DB 1; Length 1536;
 Best Local Similarity 19.4%; Pred. No. 37;
 Matches 103; Conservative 71; Mismatches 163; Indels 193; Gaps 28;

QY 58 SAPFGSSAISNGKPHQREFENLSA-----EAVTRGVGMSHTWCTSTRIHPMESL 111
 Db 1044 SGSDGSSIASRQAPYQEMSLDLHRSRYOKLR-----SNDECKVPQSPPEEP 1098
 QY 112 G-IGRPLSNDPAEDD-----KENNELYSEA 136
 Db 1099 NTIEBELIDEAEKPNLTGNLVEANSOGIIQNRISFNLFANTNIYIFPRHWTIY--- 1155
 QY 137 ERLIGTSTKEFDESIRHTLVRLS-----QDAKQRIFRP 173
 Db 1156 ERLT--EIKMNERVTEINRTSTVTFKDLDLLSSQLSEMGDFVCEYAKQVRLSR- 1212
 QY 174 LPLACHRLKNAPYVEVHSAENLFHSIYNDKQKLFLL-----LTNHRCTRLALTCGY 227
 Db 1213 -----RLINGDLEHQWF--EESLRQAYN-NKAFKLYTIDKVTQSLVKHAHT--LMTDAK 1261
 QY 228 EKKIGAAEVRNLLATRNPPSSQLDSYIMAKYVVLASGAIGNPILYNSGSLQVTPNDLS 287

1262 TABIMALFVXD-----RN-----ASTTSKQDIYR-----LQVR-SHMS 1295
 288 LIPNLGR-----YITEQPMAPCOIVLRQEFVDSVRDDPYGLPWKVAQHIANKPT 339
 1296 NTEMFRIEDKTLHVSIOYIALDDTLKEPKADEK-----WKYYVTSYALPHPT 1347
 340 DALPPIPRDPEQVTPFTTEHPHWHQIHRDAFSYGAVGPEVDSRVIVDLRWFGATDPEA 399
 1348 EGI-----PHEKKIPFLB-----RLIEGQDIDGTEVEDEP-----SPEG 1383
 400 NNL-----LVFNQNDVQG-YSM--POPTPRYRPSTASNVRRARMMADMCEVASNLG 447
 1384 ISVSTLIKIKIPIYQIHLHNGSYDVFTKATNKY-PTIANDNTQKMGVSQKKEKIS--- 1439
 448 GYLPTSPPOFMD--PGLALHLAGTTRIGDK-----ATTVAONNSL 486
 1440 -----KFLDCAVGLRNNLDEAQLSMOKWENLKDSIAKTSAGNOGI 1481

RESULT 12

AMY ASPOR STANDARD; PRT; 612 AA.
 P36914;
 01-JUN-1994 (Rel. 29, Created)
 01-JUN-1994 (Rel. 29, Last sequence update)
 30-MAY-2000 (Rel. 39, Last annotation update)
 Glucoamylase precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
 (1,4-alpha-D-glucan glucohydrolase).
 GLAA.
 Aspergillus oryzae.
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 NCBI_TaxID=5062;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=91254744; PubMed=1368680;
 Hata Y., Kitamoto K., Gomi K., Kumagai C., Tamura G., Hara S.;
 "The glucoamylase cDNA from Aspergillus oryzae: its cloning,
 nucleotide sequence, and expression in Saccharomyces cerevisiae.";
 Agric. Biol. Chem. 55:941-949(1991).
 [2]
 SEQUENCE FROM N.A.
 MEDLINE=92104497; PubMed=1761224;
 Hata Y., Tsuchiya K., Kitamoto K., Gomi K., Kumagai C., Tamura G.,
 Hara S.;
 "Nucleotide sequence and expression of the glucoamylase-encoding gene
 (glaa) from Aspergillus oryzae.";
 Gene 108:145-150(1991).
 [3]
 SEQUENCE FROM N.A.
 STRAIN=RIB 40;
 Hata Y., Tamura G., Kumagai C., Gomi K., Kitamoto K., Tsuchiya K.,
 Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.

-1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
 Glucose residues successively from non-reducing ends of the chains
 with release of beta-D-glucose.
 -1- SIMILARITY: BELONGS TO FAMILY 15 OF GLYCOSYL HYDROLASES.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; D01035; BAA00841.1; -
 EMBL; D10698; BAA01540.1; -
 PIR; J01346; J01346.
 HSP; P04064; ICAI
 InterPro; IPR002044; CBD_4.

DR InterPro; IPR000165; Glyco_hydro_15.
 DR Pfam; PF00686; CBM_20; 1.
 DR PRINTS; Glyco_hydro_15; 1.
 DR PROSITE; PRO0736; GLHYDRLASE15.
 DR PRODOM; PD001568; CBD_4; 1.
 DR PROSITE; PS00820; GLUCOAMYLASE; 1.
 KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
 Signal.
 FT SIGNAL 1 19 BY SIMILARITY.
 FT PROPEP 20 25 BY SIMILARITY.
 FT CHAIN 26 612 GLUCOAMYLASE.
 FT BINDING 146 146 SUBSTRATE (BY SIMILARITY).
 FT ACT_SITE 202 202 CATALYTIC BASE (BY SIMILARITY).
 FT ACT_SITE 205 205 GENERAL ACID CATALYST (BY SIMILARITY).
 FT ACT_SITE 206 206 INTERACT WITH SUBSTRATES (BY SIMILARITY).
 FT DISULFID 236 239 BY SIMILARITY.
 FT DISULFID 248 475 BY SIMILARITY.
 FT CARBOHYD 288 296 BY SIMILARITY.
 FT CARBOHYD 39 39 N-LINKED (GLUCNA...) (POTENTIAL).
 SQ SEQUENCE 612 AA; 65486 MW; CD7B23E5FA978F97 CRC64;
 Query Match 3.3%; Score 97; DB 1; Length 612;
 Best Local Similarity 20.5%; Pred. No. 10;
 Matches 105; Conservative 56; Mismatches 161; Indels 190; Gaps 28;
 QY 78 ENLSAEAVTRGV-----GGMSTHTWCSTPRIHPMESLPGIGRPKLSNDPAEDDKENNELY 133
 DB 107 EPISSQARIQISNPSSGALSS-----GGLGEPKENVDETATGAWGRPQ 150
 QY 134 SAEERLIGTSTKEFDESI---RHTLVLSLQDAYKDRORIFRPLPLACHRLKNAPYEW 190
 DB 151 RDGPALRATAMISFGEWLVSHTSIATDL-----VWV-----VRNLSYVAQ 194
 QY 191 HSAENLFHSIYNDKQKXLTLLTNHRTCLALTCG--YEKKIG-----AAEVRNL 239
 DB 195 YWSQSGP-DLMEVVGCTSFVAVSHR---ALVEGSSFAKTVGSSCPYCDQAQVRCY 249
 QY 240 LATRNPPSQDLSYIMAKYVVLASGAIGNPQILYNSGFSGLQVTPRNDLSIPNLGYITEQ 299
 DB 250 L-----QSFWTGSIQAN-----FCGGRSGKQIN-----TVLGSIHFD 283
 QY 300 PMAFQIVLRQEFVDSVRDDPYGLPWKVAQHIANKPTDALPIPP----- 346
 DB 284 PQATC-----DDATFQPCSAALANH--KVVTDSFRSIYAINSGRAENQAVA 328
 QY 347 -RDEPQVTTPTTEHPWH-----TQIHRDAFSYGAVGPEVDSRVIVD--LWVF--- 392
 DB 329 VGRYPEDS-----YNGNPWFLLTAAAEQLDLYQWDKIG-----SLAITDVSLEPKAL 380
 QY 393 ---GATDPEANLLVFNQDVQ-----DGYSMPQPTFRYRPSTASNVRRARMMADMCEVA 443
 DB 381 YSSAATGTYSASTTVYKDIVSAVKAYADGYVQVT--YAASTGS-----MAEQ----- 427
 QY 444 SNLGGYLPSPQPMQDGLALHLAGTTRIGFDKATTVADNNSLVNDFANLYVAG---NGT 500
 DB 428 ---YTKTDGQ-----TSARD---LTWSYAALLTANNRRNAV 458
 QY 501 IRTGFGEN-----FLLTSMCHAIKSARSINT 527
 DB 459 VPAPWGETAATSIPTASCTTSASGTYSVIT 490

RESULT 13

MOPT ARATH
 ID MOPT ARATH STANDARD; PRT; 902 AA.
 AC P93024; O64965; Q9FXI3;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Transcription factor MONOPTEROS (Auxin-responsive protein IAA24)
 DE (Auxin response factor 5).
 GN MP OR IAA24 OR ARFS OR ATIG19850 OR F6F9_8 OR F6F9.10.
 OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurooids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 [1]
 RN SEQUENCE FROM N.A., AND VARIANT ILE-796.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=98151364; PubMed=9482737;
 RA Hardtke C.S., Berlet T.;
 RT "The Arabidopsis gene MONOPTEROS encodes a transcription factor
 mediating embryo axis formation and vascular development.";
 RL EMBO J. 17:1405-1411(1998).
 [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,
 Buehler E., Chan A., Chao Q., Chen H., Cheuk R.P., Chin C.W.,
 Chung M.K., Conn L., Conway A.B., Creasy T.H., Dewar K.,
 Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin B.,
 Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
 Militscher J., Miranda M., Nguyen M., Niemman W.C., Osborne B.I.,
 Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 thaliana.";
 RL Nature 408:816-820(2000).
 [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=98004476; PubMed=9342315;
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 Goldsmith A.D., Lee J.M., Toriumi M.J., Yu G., Brooks S.Y., Chao Q.,
 Chen H., Karlin-Neumann G., Kim C.J., Lam B., Miranda M., Nguyen M.,
 Palm C.J., Shinn P., Southwick A.M., Davis R.W., Ecker J.R.,
 Theologis A.;
 RT "Full Length cDNA of gene F69.10 (GI:10086486).";
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 [4]
 RN SEQUENCE OF 13-902 FROM N.A.
 RX MEDLINE=98004476; PubMed=9342315;
 RA Kim J., Harter K., Theologis A.;
 RT "Protein-protein interactions among the Aux/IAA proteins.";
 Proc. Natl. Acad. Sci. U.S.A. 94:11786-11791(1997).
 CC -1- FUNCTION: Transcriptional regulator that modulates gene activities
 in response to auxin signals. Involved in vascular strand
 formation.
 CC -1- SUBUNIT: Capable of forming a dimer with IAA1. Interacts with
 IAA17/AXR3.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: Embryo axis, provascular tissues, procambium
 and some differentiated vascular regions of mature organs.
 CC -1- DEVELOPMENTAL STAGE: In early embryo and during organ development.
 CC -1- MISCELLANEOUS: Absence of the protein causes early embryonic
 lethality. Premature stop codons are associated with vascular
 defects.
 CC -1- SIMILARITY: Contains 1 TF-B3 domain.
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE AUX/IAA
 FAMILY.
 CC -1- CAUTION: Ref.2 sequence differs from that shown due to erroneous
 gene model prediction.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; AF037228; AAC39410.1; -;
 DR EMBL; AF037229; AAC60794.1; -;
 DR EMBL; AC007797; AAG12546.1; ALT_SEQ.
 DR EMBL; AF334716; AAG50094.1; -;
 DR EMBL; U79557; AAB92476.1; -;
 DR PIR; T51807; T51807.
 DR TRANSFAC; T04507; -;
 DR InterPro; IPR003311; AUX_1AA.
 DR InterPro; IPR003340; TF_B3.
 DR Pfam; PF02309; AUX_1AA; 1.
 DR Pfam; PF02362; B3; 1.
 KW Multigene family; Nuclear protein; Transcription regulation;
 KW DNA-binding.
 FT DOMAIN 119 306 TF-B3
 FT DOMAIN 722 902 AUX/IAA-LIKE.
 FT DOMAIN 23 28 POLY-THR.
 FT DOMAIN 282 289 POLY-ALA.
 FT DOMAIN 445 499 MET-RICH.
 FT VARIANT 796 796 T -> I.
 SQ SEQUENCE 902 AA; 99650 MW; D20D348369757F09 CRC64;
 Query Match 3.3%; Score 96.5; DB 1; Length 902;
 Best Local Similarity 18.8%; Pred. No. 20;
 Matches 83; Conservative 63; Mismatches 133; Indels 163; Gaps 21;
 QY 14 YHKKEIEFQKIDIRFNVKIG-----ALQGVSVPRNVNPTLDPGA--- 56
 DB 459 HNNQNVPSFSEMQQ--NIVMGNGLLGDMKMQOPLMMKQKSEWQPNKLTWNPSASNT 515
 QY 57 -----WSAP--PGSSAI---SNGKNPQREFENLSAEAVTRGVGGMSTHTCTSP 101
 DB 516 SGQEQNLQSMSAPAKPENSTLSGSSGRVQHGLESMEQASQVT-----ISTVCNEE 568
 QY 102 RIHPMESLPGIGRPKLSNDPAEDDK---EWNELYS-EAERLIGTSTKEFDEIRHTLVL 157
 DB 569 KVNQLQK-PG-----ASSPVQADQCLDITHIQYQSPDPINGFSLFDELTSQVSSP 621
 QY 158 RSLQDAYK-----DRQRIFRPLPLACHRLKNAPEYVEMHSAENLFHSIYND-- 204
 DB 622 QSLAGSYKQPFILSSQDSSAVLP-----DSTNSP-----LFHDVWDITQLNGLK 665
 QY 205 -----KQKLFLLTNRCTRLALTGGYKKGAAEVRNLLATRNPSQLDSYIMAKV 257
 DB 666 FDFQSPLMQDLYA--SQNICMSNTTS-----NILDPLSLNTVLDQFCAIK- 710
 QY 258 YVLASGAIGNPQILYNSGFGSLQVTRNDSLPNLGRYTEQPMACQIVLROEFVD--- 314
 DB 711 -----DTDFQNHPSGGLVGNNTSFADQVQSQITSASFADSQAFSRQDFPDNSG 759
 QY 315 -----SVRDDPYGLPWKEAQAQIAKNPDTALPIPRDPSPQVTTPTTEE 360
 DB 760 GTGTSSNVDFDCSLRQNSGSSWOKIAT-----PRVKT----- 794
 QY 361 HPWHTQIHRDAFSYGAVGPEVD 382
 DB 795 ---YTKVQKT---GSVGRSID 809
 RESULT 14
 US02 HSV11
 ID US02 HSV11 STANDARD; PRT; 291 AA.
 AC P06435;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Protein US2.
 GN US2.
 OS Herpes simplex virus (type 1 / strain 17).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

Db 783 LDATDDVIRNDG-DIPTNSAFDFSSKSNASN 814

Search completed: October 28, 2003, 15:57:59
Job time : 32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.
OM protein - protein search, using sw model
Run on: October 28, 2003, 15:47:46 ; Search time 103 Seconds
(without alignments)
1360.413 Million cell updates/sec

Title: US-09-856-327-2_COPY_76_618
Perfect score: 2901
Sequence: 1 NAEETAVPVPGYHKKNEI.....IINTLGGTDGKNTGHRNL 543

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 23:*
1: sp arChaea:*
2: sp bacteria:*
3: sp fungi:*
4: sp human:*
5: sp invertebrate:*
6: sp mammal:*
7: sp mhc:*
8: sp organelle:*
9: sp phage:*
10: sp plant:*
11: sp rodent:*
12: sp virus:*
13: sp vertebrate:*
14: sp unclassified:*
15: sp virus:*
16: sp bacteriap:*
17: sp archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	1156.5	39.9	623	3 Q8J136	Q8J136 peniophora
2	976.5	33.7	564	3 Q8J2V8	Q8J2V8 tricholoma
3	175.5	6.0	522	2 Q8GAK8	Q8GAK8 arthrobaete
4	173.5	6.0	539	2 Q8QEE7	Q8QEE7 burkholderi
5	170.5	5.9	573	16 Q9P190	Q9P190 campylobact
6	170	5.9	551	2 Q9RH54	Q9RH54 pantoea agg
7	166.5	5.7	579	16 Q9A7T6	Q9A7T6 caulobacter
8	159	5.5	615	2 Q34214	Q34214 pectobacter
9	158.5	5.5	529	17 Q9HOR8	Q9HOR8 halobacteri
10	158.5	5.5	619	16 Q8YMW9	Q8YMW9 anabaena sp
11	157	5.4	545	2 Q9LBR8	Q9LBR8 gluconobact
12	155	5.3	539	16 Q8YLB9	Q8YLB9 ralstonia s
13	148	5.1	553	2 Q9XCR0	Q9XCR0 pantoea cit
14	144	5.0	722	16 Q9R226	Q9R226 deinetococcus
15	141	4.9	523	16 Q98C76	Q98C76 rhizobium 1
16	141	4.9	591	16 Q911K8	Q911K8 pseudomonas

17	135	4.7	561	16 Q8U7S2	Q8U7S2 agrobacteri
18	131.5	4.5	528	2 Q8KPV4	Q8KPV4 synechococc
19	128.5	4.4	748	10 Q9M0H4	Q9M0H4 arabidopsis
20	128.5	4.4	748	10 Q8LDP0	Q8LDP0 arabidopsis
21	128.5	4.4	748	10 Q94BP3	Q94BP3 arabidopsis
22	127	4.4	499	16 Q988P2	Q988P2 rhizobium 1
23	123.5	4.3	508	16 Q8PKP1	Q8PKP1 xanthomonas
24	122	4.2	527	16 Q8U803	Q8U803 agrobacteri
25	119	4.1	518	16 Q8EYN5	Q8EYN5 leptospira
26	117.5	4.1	620	3 Q12623	Q12623 humicola gr
27	116.5	4.0	494	16 Q8FY45	Q8FY45 brucella su
28	116.5	4.0	609	2 Q8KIL0	Q8KIL0 rhizobium e
29	116	4.0	502	5 Q18672	Q18672 caenorhabdi
30	116	4.0	599	5 Q18429	Q18429 caenorhabdi
31	114.5	3.9	704	3 Q9P8D9	Q9P8D9 candida tro
32	114	3.9	769	3 Q74253	Q74253 pycnoporus
33	113	3.9	1012	10 Q64492	Q64492 arabidopsis
34	113.5	3.9	544	16 Q8XU48	Q8XU48 ralstonia s
35	113.5	3.9	578	16 Q57307	Q57307 mycobacteri
36	112	3.9	768	3 Q8J2T4	Q8J2T4 grifola fro
37	112	3.9	1881	12 Q9DUN3	Q9DUN3 vesicular e
38	111.5	3.8	494	16 Q8YJ07	Q8YJ07 brucella me
39	111.5	3.8	586	10 Q8H094	Q8H094 oryza sativ
40	111	3.8	1492	5 Q8SSU1	Q8SSU1 dictyosteli
41	110.5	3.8	589	10 Q9M4V5	Q9M4V5 oryza sativ
42	110	3.8	537	16 Q8F9Z5	Q8F9Z5 leptospira
43	110	3.8	551	2 Q8GME7	Q8GME7 streptomyce
44	110	3.8	1498	16 Q8YAQ5	Q8YAQ5 listeria mo
45	109	3.8	418	16 Q8EL08	Q8EL08 oceanobacil

ALIGNMENTS

RESULT 1

Q8J136 PRELIMINARY; PRT; 623 AA.
AC Q8J136
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Pyranose oxidase (EC 1.1.3.10).
OS Peniophora sp. SG.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Aphyllophorales; Lachnocladiaceae; Peniophora.
OX NCBI_TaxID=204723;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SG;
RA Heckmann D.M., Bastian S., Rekowski M., Giffhorn F.;
RT "Pyranose Oxidase of the White-Rot Fungus Peniophora sp. Strain SG:
RT Cloning and Characterization of the Gene, Heterologous Expression in
RT Escherichia coli, and Properties of the Recombinant Enzyme."
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF535193; AAO13382.1;
KW Oxidoreductase.
SQ SEQUENCE 623 AA; 69342 MW; 4B7DB64A387F43D5 CRC64;

Query Match	39.9%	Score 1156.5;	DB 3;	Length 623;
Best Local Similarity	45.7%;	Pred. No. 7.3e-89;		
Matches	253;	Conservative 69;	Mismatches 159;	Indels 73; Gaps 15;
Oy	13	GYHKNEIEFKQIDRFVWIKGALOOVSVPRNQNVPTLDPGWSAPPGSSAISNGKNP	72	
Db	87	GAHKNTVEQKIDKFKVVIQQQLMSVSPVNTLVVDTLSPTSQA--STFFVRNGSNP	144	
Oy	73	HOREFENLSAEATRVGGGMSHTWCTSPRIHPMESLPGIGRPKLSNDPAE-DDKEWNE	131	
Db	145	EQDPLRNLSCQAVTRVVGNGSTHTWCATPFDREQ-----RPLLVKDDADDAEDWR	197	
Oy	132	LYSEAEERLIGTSTKEFDESIRHTLVRSLODAYKORIFRPLPLACHRLKNAPEYVEWH	191	
Db	198	LYTKAESYFOTGTDQPKESIRHNLVNLKLAEEYKQ-QRDFQOIPLAATR--RSPTTFVWS	254	

QY 192 SAENLF--HSTYNDKOKKFTLLTNHRTALTGVEKKIGAEVRNLLATNPSSQL 249
 DB 255 SANTVFDLQNRPTDAEERNLFPVACERV-----VRNAL-----NSEI 295
 QY 250 DSY-----IMAKVYVLASGAIGNPQILYNSFGSL-QVTPRN-DSLIPNLGRY 295
 DB 296 ESLHIDLISGRPEIKADVVVLTAGAVHNTQLLVNSFGQLGRPNPTNPPELLPSLSGY 355
 QY 296 ITEQPMAPCQIVLRQEFVDSVRDPP--YGLP-----WKEAQAQH 333
 DB 356 ITEQSLVFCQVMSTELISVKSMTIRGTPGELTYSVTYTPGASTNKHKPDWNEKVKNH 415
 QY 334 IAKNPTDALPFPDRPDPQVTPTEHPHWHQIHRDAFSGVAGPPEVDSRVLDLRFWG 393
 DB 416 MMQOEDEPLIPEDPDPQVTTLPQSPHWHQIHRDAFSGVAGVQOISDRLLVDWRFPG 475
 QY 394 ATDPEANLLVFNQDVODGYSMPQPTFRYR-PTASNVRRARKMADMCVSNLGGYLPT 452
 DB 476 RTEPKENKLFWSKIIDAYNMPOPTDFRFPAGRTSKEAEDMTDMCVMSAKIGGFLPG 535
 QY 453 SPQFMDPGLALHLAGTTRGFQKA--TTVADNNSLVWDFANLYVAGNGTIRTGGENPT 510
 DB 536 SLPOFMEPGLVLLGGTHRMGFEKEDNCCVNTDSRVFGFNLFGLGCGNIPYAGANPT 595
 QY 511 LTSMCHAIKSARSI 524
 DB 596 LTMSLAIKSCEYI 609
 RESULT 2
 Q8J2V8
 ID Q8J2V8 PRELIMINARY; PRT; 564 AA.
 AC Q8J2V8;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Pyranose 2-oxidase (EC 1.1.3.10).
 GN P20.
 OS Tricholoma matsutake.
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 OC Agaricales; Tricholomataceae; Tricholoma.
 CX NCBI_TaxID=40145;
 RN [1]
 RP Takakura Y.;
 RA "Purification, characterization and primary structure of a pyranose
 RT oxidase from Tricholoma matsutake."
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB043883; BAC24805.1;
 KW Oxidoreductase;
 SQ SEQUENCE 564 AA; 61942 MW; 134790030045FCIB CRC64;
 Query Match 33.7%; Score 976.5; DB 3; Length 564;
 Best Local Similarity 41.3%; Pred. No. 1e-73;
 Matches 223; Conservative 82; Mismatches 182; Indels 53; Gaps 13;
 QY 2 AEGTAVPYPGYHKKNEISFQKIDRPNVNIKALQOVSVVRNQNVTPLDPCGANSAPP 61
 DB 62 AEIGSQDNPIVGAHRHSIKFQDKTFVNIINGALQIPISISDITYQPTLVAVAAM-APP 120
 QY 62 GSSA----ISNGKNPHOREFENISAEAVTRGVGSMTHWTCSPRIHPHWPESLPGIGRPK 117
 DB 121 IDPAEGQLVINGHNPQAEGLNLPFGSAVTRVGVGMATHWTCACPTPHD-----EE 170
 QY 118 LSNDAEDDKEMNELYSEAEIRLTGTKEPDESIRHTLVLSQDAYKDRQIRFRPLPLA 177
 DB 171 RVNPNV-DKQEPDALLERAKTLNVHSDQYDDSIQIVVKETLQOQL-DASRGVTLPLG 228
 QY 178 CHLKNAPVEVHSHENLPHSYNDKQKKLFTLLTNHRTALTGVEKKIGAEVR 237
 DB 229 VERTDNPIVTVTGTGAD----TVLGDVPKSPREVLVTETRVTKFIVSETNPTQVVAALLR 284

QY 238 NLLATNPSSQLDSYIMAKVYVLASGAIGNPQILYNSFGSLQVTPRNDLSIENLGRYIT 297
 DB 285 NL-----NTSN--DELVAQSFVIACGAVCTPQILWNS-----NIRH-----ALGRYLS 327
 QY 298 EQPMAPCQIVLRQEFVDSVRDPPYGLPMMKEAQAHIKNPTDALPFPDRPDPQVTPPF 357
 DB 328 EQSMTFCQIVLKRISIVDSIATDPR-----FAAKVEAHKKKHPDDVLPIPFHEPSPQVMIPY 383
 QY 358 TSEHPHWHQIHRDAFSGVAGPPEVDSRVLDLRFWGATDPEANLLVFP-----QND 408
 DB 384 TSDFPWHVQVHR--YAFDVGPKADPRVVVDLRFQKSDIVEENRVTFGNPKLRDWEAG 441
 QY 409 VODGYSMPQPTFRYRPTASNVRRARKMADMCVSNLGGYLPTSPQFMDPGLALHLAG 469
 DB 442 VDTYGMPOPTFHVKTNDAGDQDMMDMTNVANILGGYLPSPQFMAPGLAQHITG 501
 QY 469 TTRIGEDKATTVADNNSLVWDFANLYVAGNGTIRTGGENPTLTSCHAIKSARSIINTL 528
 DB 502 TTRIGTDDQTSVADPTSKVHNFNVLWVGNGCIPDATACTNPTRTSVAYALKGAEVVSYL 561
 RESULT 3
 Q8GAK8
 ID Q8GAK8 PRELIMINARY; PRT; 522 AA.
 AC Q8GAK8;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE 2-keto-glucuronate dehydrogenase-like protein.
 GN KDG.
 OS Arthrobacter nicotinovorans.
 OC Plasmid PA01.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Micrococcineae; Micrococcaceae; Arthrobacter.
 CX NCBI_TaxID=29320;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95115562; PubMed=7815950;
 RA Grether-Beck S., Igloi G.L., Pust S., Schiltz E., Decker K.,
 RA Brandsch R.;
 RT "Structural analysis and molybdenum-dependent expression of the PA01-
 RL encoded nicotine dehydrogenase genes of Arthrobacter nicotinovorans."
 RL Mol. Microbiol. 13:929-936(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96172783; PubMed=8588735;
 RA Menendez C., Igloi G., Henninger H., Brandsch R.;
 RT "A PA01-encoded molybdopterine cofactor gene (moaA) of Arthrobacter
 RT nicotinovorans: characterization and site-directed mutagenesis of the
 RT encoded protein."
 RL Arch. Microbiol. 164:142-151(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97230479; PubMed=9073590;
 RA Menendez C., Igloi G.L., Brandsch R.;
 RT "IS1473, a putative insertion sequence identified in the plasmid PA01
 RT from Arthrobacter nicotinovorans: isolation, characterisation and
 RL distribution among Arthrobacter species."
 RL Plasmid 37:35-41(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98088982; PubMed=9428706;
 RA Menendez C., Otto A., Igloi G., Nick P., Brandsch R., Schubach B.,
 RA Botcher B., Brandsch R.;
 RT "Molybdate-uptake genes and molybdopterine-biosynthesis genes on a
 RT bacterial plasmid. Characterization of MoaA as a filament-forming
 RT protein with adenosinetriphosphatase activity."
 RL Eur. J. Biochem. 250:524-531(1997).
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Schenk S., Hoelz A., Kraus B., Decker K.;
 RT "Gene structure and properties of enzymes of the plasmid-encoded
 RT nicotine catabolism of Arthrobacter nicotinovorans."

J. Mol. Biol. 284:1323-1339(1999).
[6]
SEQUENCE FROM N.A.
MEDLINE=21405725; PubMed=11514508;
Baitsch D., Sandu C., Brandesch R., Igloi G.L.;
RT "A gene cluster on pAOI of *Arthrobacter nicotinovorans* involved in the
RT degradation of the plant alkaloid nicotine: cloning, purification and
RT characterization of 2,6-dihydroxypyridine 3-hydroxylase".
J. Bacteriol. 183:5262-5267(2001).
[7]
SEQUENCE FROM N.A.
Igloi G.L., Brandesch R.;
RT "Sequence of the 165 kb Catabolic Plasmid pAOI from *Arthrobacter*
RT *nicotinovorans* and Identification of a pAOI-dependent Nicotine Uptake
RT System".
JL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
JL EMBL: AJ507836; CAD47896.1; -;
JL Plasmid.
JL SEQUENCE 522 AA; 56526 MW; ADB9B35DB3CCC939 CRC64;
Query Match 6.0%; Score 175.5; DB 2; Length 522;
Best Local Similarity 21.8%; Pred. No. 5.2e-06;
Matches 121; Conservative 58; Mismatches 208; Indels 167; Gaps 24;
Y 27 DRFVNVIKALQOVSVVRNQNVTLPDCAWAP---PGSSAISNGKNPHOREFENLSAE 83
b 47 DXFINDERESYKLT-----WTDTRLATGWSLAKDFFGSPA-WNGK----- 87
Y 84 AVTRGVGGMSTHTWCSTPRI-----HPPMESLPQIGRPLKNSDPAEDDKENNELYSEA 136
b 88 ---AVGGTATFWTGLTPFKWHEFKTHYYGLPD---GTIADWPLDLD-ELDHYYTAA 139
Y 137 ERLICTSKPEDESHTVLVLSLODAYKQRIIFRPLPLACHRLKNAPEYVEM-HSAEN 195
b 140 EKAVGASH-----RHGRPELPASNGYK-----VLANGAERIGYRHYATG 178
Y 196 LFHS-----IYNDKQKXKLLTLLTHNRCRLALTGGEKKIGAEVRNL 239
b 179 PYATNVEPYDGRPGTVQDGFAMAGDSRAKWSPLVSEIPKALA-TGLLELRTAQAVQ-- 235
Y 240 LATRNPPSQLDSYI-----MAKVVYLASGAIGNPQILYNSGSLQVTPRNDSL 288
b 236 -ITLPGDGRAGVVYTDVGNLQGRARLVAVAGNAIETPRLLLSATSG-----HTDGL 289
Y 289 -----IPNLGYITEQPMFQCOIVLRQBFVDSVRDDP---YGL 323
b 290 ANSSGOVGRNMYRHTTGLVYAEFFNEVHMVYRGEPMAGI-----ISDSRHDPGRGFGV 342
Y 324 PWKEAQAHIKNTDAPLPIPRDPEPQVTPTEHPHWTQIHRDAFSYGAVGPEVDS 383
b 343 GYIEMIAQGLPSFT-----FMSF-----GEWGPQFTE 371
Y 384 RVIVDLR-----WFGATD-PEANLLLVFONDVQDGYSMPOQTFYRPPSTASNVRAKQWAD 438
b 372 KVEAYTRTAALWICGEDVPOASNVRLSTNTVDRGLGPAPVHY-DHPNDIAVRNHGYE 430
Y 439 MCEVASNLGGVLPSPQFMDPGALHLAGTTRIGFKATTVDNNSLVDFANLYVAGN 498
b 431 QGEKLFKSVGAIRTRAPGMSG---HNLGTARMSQKPDGCVNSFGQAHVDPNLFVSDG 487
Y 499 GTIRTFGENPTLT 512
b 488 SOFTTGAANPTLT 501

RESULT 4

Q9GQ7 PRELIMINARY; PRT; 539 AA.
Q9GQ7;
01-MAR-2003 (TrEMBLrel. 23, Created)
01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
Glucose dehydrogenase.

GN GDHALPHA.
OS Burkholderia cepacia (pseudomonas cepacia).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=292;
RN [1]
RP SEQUENCE FROM N.A.
RA Sode K., Inose K., Fujikawa M., Yamazaki T.;
RT "Molecular cloning of the gene encoding thermostable Glucose
RT dehydrogenase from a moderate thermophilic bacterium".
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF430844; AAN39686.1; -;
SQ SEQUENCE 539 AA; 59832 MW; 9CB4A4E5D39AE845 CRC64;

Query Match 6.0%; Score 173.5; DB 2; Length 539;
Best Local Similarity 19.9%; Pred. No. 8e-06;
Matches 109; Conservative 80; Mismatches 191; Indels 169; Gaps 24;

QY 60 PGSSAISNGKNPHOREFENLSAEAVTRGVGGMSTHTWCSTPRIHP---PMESLPQIGRP 116
b 78 PNDVYLILKGEHKFSQV-----IRAVGGTTWHWAASAMRFIPNDFKMSVYGVGR- 128
QY 117 KLSNDPAE-DDKENNELYSEAEERLIGTSTKEPDESIRHTLVLSLODAYKQRIIP--RP 173
b 129 ---DWPIQVDDLE--PYQRAEEELGVMPGPEE-----DLYSPKQPYMPMP 171
QY 174 LPLACHR--LKNAPYEVVHSAENLFHSI-----YNDKQKXKLLTLLTHNRCRLALTOG 226
b 172 LPLSFNEQITKA-----LNNYDPKFXVTVBPVARNRPYDGRPTCCGNNNCMPICPIGA 226
QY 227 Y-----BKKGAAEVRNLATR---NFSQL-----DSYIMAKVYVVLASG 263
b 227 MVNGIVHVEKAERAGAKLIENAVVYKLETPDKRIVAALYKDKTGAHRVEGKYFVLAAN 286
QY 264 AIGNPQILYNSGSLQVTPRNDSLIPNLGYITEQPMFQCOIVLRQBFVDSVRDDPYGL 323
b 287 GIETPKIL-----LMSANRDFPNGV 306
QY 324 PWKEAQAHIKNTDAPLPI-----PFRDPEPQVTPTEHPHWT-----QIH--- 368
b 307 ANSDMVGRNLMHPGTVGSYASEKLMWPGRQPMETSLIGFRDGPFRATEAAKKIHLSN 366
QY 369 -----RDAPSYGAV--GPEVDSRV-----TVDLRFQATDPEANLLLVFONDVODG 412
b 367 LSRIDQETQKIFKAGKMKPDELDAQIRDSARYVQDFHEILPQENRIVPSKTATDA 426
QY 413 YSMPOQTFYRPPSTASNVRAKQWADMEV-----ASNLGG-----YLPSPQFMDP 460
b 427 IGIPRPEITY-----AIDDYVKRGAHTREVYATAAKVLGGTDVVVDFNPFNN----- 475
QY 461 GLALHLAGTTRIGFKATTVDNNSLVDFANLYVAGNRTGTGFGENTLTSCHAIKS 520
b 476 ---HITGSTMIGADARDSVVDKCRTHPHNLFSSATMPTVGTNVTLTIALALRM 531
QY 521 ARSIINTLK 529
b 532 S---DTLK 536

RESULT 5

Q9PI90 PRELIMINARY; PRT; 573 AA.
Q9PI90;
01-OCT-2000 (TrEMBLrel. 15, Created)
01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
Putative oxidoreductase subunit.
GN CJ0415.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=NCTC 11168;
 RX MEDLINE=20150912; PubMed=10688204;
 RA Parkhill J., Wren B.W., Mungall K., Kelsey J.M., Churcher C.,
 RA Basham D., Chillingworth T., Davies R.W., Feltwell T., Holtroyd S.,
 RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
 RA Whitehead S., Barrell B.G.;
 RT "The genome sequence of the food-borne pathogen *Campylobacter jejuni*
 RT reveals hypervariable sequences.";
 RL Nature 403:668-668(2000).
 DR EMBL; ALJ39075; CAB74251.1; -;
 KW Complete proteome.
 SQ SEQUENCE 573 AA; 63686 MW; 7BEB522A779A5F94 CRC64;

Query Match 5.9%; Score 170.5; DB 16; Length 573;
 Best Local Similarity 20.9%; Pred. No. 1.6e-05;
 Matches 111; Conservative 61; Mismatches 237; Indels 123; Gaps 21;

QY 77 FENLSAEAVTRGVGGMSTHTCSTPRTHPPWESLPGIGRP-----KLSNDPAEDD-----K 127
 DB 84 YRMGSLGNNVGGAGVHNGWTFRPMYPDFPIQLSKORYGNKLGNDYTLQDWGVYK 143
 QY 128 EWNELYSAEERLIGTSTKEP-----DESIHRTLVLRSLODAYKQRIER 172
 DB 144 DMEPYDRFEKTCGVSGEPNLAEMGAFRSSYPQEPLENTKMLKRFESAASSNLHTY 203
 QY 173 PLPLA-----CHRLKNAPEYVEHSAENLFHSIYNDKQKLF 210
 DB 204 RLPASNSKGGYTNPDGODLAPCOYCAYCERFG-----CEYGAKASPLNTVIPKAMSTGKY 258
 QY 211 TLTLNHRCTRLALTGGYKKGAEVNRNLLATRNPSQLDSYIM-AKVYVLASGAIGNPQ 269
 DB 259 TITYSNVTLQ-----KKQKVTGVKFDVTRT-----MKEYIQADIVVLTSYMFENNAK 308
 QY 270 ILYNSGFGSLQVTP-----RNDSLIPNLG--RYITEQPMACQIVLRQEFVDSVRDD 319
 DB 309 LLAVSNI-GRQYDPKTKGTLGRNYCYQNMGTATAFDEQFNTF-----MSSGALGTTSD 363
 QY 320 PYG--LPWKE-----AVAQHIKNTDALLIPFRDPPEQVTPPTTEHPHWTQIHRD 370
 DB 364 FNGDNFDSHKEFLHGAMIYSVOLGTRPIQSAFLPAGAP-----TWGAEF-KK 410
 QY 371 AFSVG-----AVGPEVDSRVIVDLRFEGATDPEANLLVFQNDVQDGYSMQPTFYRPS 425
 DB 411 ALNTNFRATVGGQ-----GASLPHKNYLSLDPTYKDAFGMLRLTYN-F 457
 QY 426 TASNVRARKMAD-MCEVASNLGGLYPTSPQFMD-----PGLALHLAGTTRIGFDKATT 479
 DB 458 TDQDRALHKEITDKTAEAVKRMQGVKSIKKGAVLKDYVVYPQYSTHTNGTGTGADRETS 517
 QY 480 VADNLSLVDFANLYVAGNGTIRTGFGENTLTSMCHAIKSARSINTLKG 531
 DB 518 VVNTYLOHMDADNLFFVVGAGNFQHNSGYNPTDTVGTALAYRCAEGILKYHSG 569

RESULT 6

Q9RH54 PRELIMINARY; PRT; 551 AA.
 AC Q9RH54;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Dehydrogenase subunit 1.
 OS Pantoea agglomerans.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Pantoea.
 CX NCBI_TaxID=549;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chang K.-I., Yum D.-Y., Pan J.-G., Shin Y.-C.;
 RT "Cloning and expression of a gene cluster encoding three subunits of
 RT membrane-bound 2-keto-D-gluconate dehydrogenase from *Erwinia herbicola*

RT ATCC08111 in *Escherichia coli*;"
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF068066; AAF21261.1; -;
 DR InterPro; IPR000205; NAD binding.
 SQ SEQUENCE 551 AA; 60047 MW; FE2ECCD0933DAFBF CRC64;

Query Match 5.9%; Score 170; DB 2; Length 551;
 Best Local Similarity 20.3%; Pred. No. 1.6e-05;
 Matches 109; Conservative 65; Mismatches 209; Indels 154; Gaps 21;

QY 68 NGKNPHOREFENLSAEAVTRGVGGMSTHTCSTPRTHPP-----MESLPGIGRPLKSLNDPAE 124
 DB 93 NVTGPNADSFQ-----QGYLRTVGGTTHWAASCRHHPSPDFVMSQYKGVGR----- 139
 QY 125 DDKEM-----NEL-----YSEAEERLIGTSTKEFDESIRHTLVLSLODAYKQRIERPLA 177
 DB 140 ---DWPIGVDELEPWYCKAENEIGVAGP--NDPARQSPTERS-----QPYPM- 181
 QY 178 CHRLKNAPEYVEHSAENLFHSIYNDKQKLF-----FTLLNHRCTRLAL 223
 DB 182 -----DMVPPAHGDNYPASVNPVPHGYNLVPVPOGRSTRPWEGRPTCCGNNNCQFICP 233
 QY 224 TGYEYKIGAAEVRNLLATRNPSQLDSYIM-----AK 256
 DB 234 IGAMTYNGIHHVE-----RAERNGAVVLAEAVVYKMDTDSNNRITAVHMLDTSGASHKATAK 289
 QY 257 VVVLASGALGNQILYNSGFGSLQVTPRNDSLIPN---LGRVITEQPMACQIVLRQEF 312
 DB 290 AFALACNGIETPRLLD-----MAANDANPNGIANASDMVGRNMDSHGFCFSFLTKE-- 341
 QY 313 VDSVRDDPYGLPWKEAQAHIKNT-DALPIPRDPPEQVTPPTTEHPHWTQI----- 367
 DB 342 -----PW-----LCKGPAQSSCMVGYRDG-----FRDYSANKVILNNI 377
 QY 368 -----HRDAFSGVAGVPEVDSRV-----IVDLRFEGATDPEANLLVFQNDVQDGYSM 415
 DB 378 SRVVTATQAMKGLVGLKALDEIRYRAVHSVDLSISLEPLDPENRLTLSTRKDPHGL 437
 QY 416 PPTFYRSTASNVKARKMADCEVASNLGGLYPTSPQFMDPGLAL--HLAGTTRIG 473
 DB 438 PCPDIYDVGVYRKGAESAHAQLEHI-----GQLFDAKEFTISQGLNANNHMGVING 492
 QY 474 FDKATTVADNLSLVDFANLYVAGNGTIRTGFGENTLTSMCHAIKSARSINTLKG 530
 DB 493 KNAKEAVDVGNCRAFDHENLWLPGGGAI PSASVVNSTLTMAALGLKAAHDI SLRMKG 549

RESULT 7

Q9A7T6 PRELIMINARY; PRT; 579 AA.
 AC Q9A7T6;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JAN-2002 (TrEMBLrel. 20, Last annotation update)
 DE Oxidoreductase, GMC family.
 GN CC1634.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
 OC Caulobacteraceae; Caulobacter.
 CX NCBI_TaxID=155892;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=21173698; PubMed=11259647;
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
 RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RT "Complete genome sequence of *Caulobacter crescentus*."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).

DR ENBL; AB005839; AAK23612.1; --
 DR TIGR; CC1634; --
 DR InterPro; IPR000205; NAD_binding.
 KW Complete proteome.
 SQ SEQUENCE 579 AA; 64397 MW; 7125C4DPAD618F10 CRC64;

Query Match 5.7%; Score 166.5; DB 16; Length 579;
 Best Local Similarity 21.2%; Pred. No. 3.5e-05;
 Matches 105; Conservative 66; Mismatches 234; Indels 91; Gaps 19;
 2y 89 VGGMSTHTWCTSPRIHPHMSLPGIGRPKLSND---PAEDDKENNELYSEAEELIGTSTK 145
 120 VGGSLTWGRQSYR-HSPID-PEANAREGIADWPRIYEDLAPW---YEHVERFIGVSGQ 174
 146 EFDESIRH---TLVLSRLQADYKORQRI-FRPLPLACHLKNAPYEVWESA--- 193
 175 A-EGLPHPFDGHYQPPMELNCEKAFKARSEARFPERRVTIGRTAHLTDPTBEQALGR 232
 194 -----ENL-----FHSIYNDKQKLFLLTNHRCRLALATGG-----YEKIGAAE 235
 233 TKQYRNLCIRGCPFGAYSSNSG---GLIAAERTGNLVIRPNSIVTELLIYDERAGRAS 288
 236 VRLNLAIRNPSSQLDSYIMAKVYVLASGAINPQILYNSGFSGLQVTPRNDSLIPN--- 291
 289 GVRIL---DAETRKDEEFHADVIFLCASALNSAWIMNS-----TSSRFNPGFCN 335
 292 ---LGRVITEQPMACQIVLRQEFVDSV-----RDDPYGLPWWKEAVAQHIANKPTDALP 343
 336 ASDQGRNVMDHHLGAGATGQAPFADMYFSGRRPNIGYVPRFRN-LGDAASKRSQDYLGR 394
 344 IPFRDPEPOVTPPTEEHHPWHQIHRDAPSYGAV-----GPEVDSRVIVDLRWFEGAT 395
 395 FGQGGAGAT-----WEDRGQGGRGFGAARKAALSQGPWT-----MGLSGFGEM 441
 396 DPEANLLVFQNDVODGVSMPQPTFRYRSTASNVRAKQWADMCEVASNLGGYLTPSP 455
 442 LPYADNRVTLNRDVEDKFGLPPLTMVNTMRDNEMAMRRDMQAAAAEMLEAAGFQNVRAHD 501
 456 QFMDPGLALHLAGTTRIGFDKATTVADNNLSLWDFANLYVAGNGTIRGFGENPLTSMC 515
 502 NGFAPGLGIHEMGTAIRMGDRPKTSVLNAHQVHECKNVYVTDGAAMASACVNPSLTYMA 561
 516 HAIKSARSIINTLKG 531
 562 LTARAADHAVRARKG 577

RESULT 8
 34214

D O34214 PRELIMINARY; PRT; 615 AA.
 C O34214;
 T 01-JAN-1998 (TrEMBLrel. 05, Created)
 T 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 T 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
 E Dehydrogenase subunit precursor.
 S Pectobacterium cyripedii.
 C Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 C Enterobacteriaceae; Pectobacterium.
 X NCBI_TaxID=55209;
 X [1]
 P SEQUENCE FROM N.A.
 C STRAIN=ATCC29267;
 C MEDLINE=98012950; PubMed=9352801;
 X Yum D.-Y., Lee Y.-P., Pan J.-G.;
 T "Cloning and expression of a gene cluster encoding three subunits of
 T membrane-bound gluconate dehydrogenase from *Erwinia cyripedii* ATCC
 T 29267 in *Escherichia coli*."
 T J. Bacteriol. 179:6566-6572 (1997).
 X ENBL; U97665; AAC45885.1; --
 X SIGNAL.
 X CHAIN
 T 1 22 DEHYDROGENASE SUBUNIT.
 T 23 615
 T SEQUENCE 615 AA; 67241 MW; B9E1A84FD035609A CRC64;

Query Match 5.5%; Score 159; DB 2; Length 615;
 Best Local Similarity 19.9%; Pred. No. 0.00017;
 Matches 124; Conservative 85; Mismatches 257; Indels 158; Gaps 32;
 QY 4 EGTAVPYVPGYHKONEIEFQKIDIRFVNI-----KGALQOVSVVRNONVPTLDPGAMS 58
 DB 5 ERVSUP-VSGYRGEGVTVANLKKVDAVVVGFGWAGAIMAKELTEAGLNVALERG--- 60
 QY 59 APFGSAISNGKNP-----HQREFENLSABAVT-----R 87
 DB 61 --PHRDTYDGAYPOSIDELTYNIRKKLFQDLKSXTVTIRHDASQTAVPYRQLAALPCT 118
 QY 88 GVGGMSTHTWCTSPRIHPHMSLPG-----IGRPKLSNDPAEDD---KENNEL---YSEAE 137
 DB 119 GTGGAGLHSGVHFVRDPVELNLRSHYEARYKNFIPEGMTIQDFGVSYNELEPFDDQAE 178
 QY 138 RLIGTSTKEFDESIRHTLVLSRLQADY--KORQIRFRPLP-----LACHRLKNAPYEVW 190
 DB 179 KVFGISGAW--TIKGMIGKEKGNFYAPDRSSD-PLPAOKRTYSQALFAQAAESVGY 235
 QY 191 H-----SAENLPHSIYND-----KQKLF 212
 DB 236 HPYDMP SANTSGPYNTYGAQMGPCNFCGCGSYACYMYSKASPNVNTLPALRQPKFEL 295
 QY 213 LTNHRCRLALATGGYEKKIGA-----AEVRNLLATRNPSQ---LDSYIMAKVYVLASGAI 266
 DB 296 RNNAYVLRLNTGDKRATGVTYLDGQGREVV---QPADLVLSAFQPHNVHMLLSGIG 352
 QY 267 NP-QILYNSGSLQVTPRNDLSLIPNL-GRVITEQPM---AFQIVLRQEFVDSVRDDPY 321
 DB 333 QPYNDITNEGVVGRNFAYQNI STLKALFDKNTTNPFFIGAGGAGVAVDDFNADNFDHGPY 412
 QY 322 GL-----PWWKEAVAQHIANKPTDALPIPRDPEPOVTPPTEEHHPWHQIHRDAPSYGAV 377
 DB 413 GFVGSPW-----VNQACTKPVSGLPTEKFTN-----WGSQWK-----RAV 450
 QY 378 GPEVDSRVIVDLRWFEGATDPEANLLVFQNDVODGYSMP---OPTFRYRSTASNVRAK 434
 DB 451 ADTNNHHISMDAH--GAHQSYRANVYLDLDPNVKNVYQGPQLLRMTFMDQD---NDIRMAQF 505
 QY 435 MNADCEV--ASN-----LGGYLTPSPQPMFPGGL--ALHLAGTTRIGFDKATTVADNNLS 486
 DB 506 MVGKRNKITEAMNPKMIIGG--AKGPGTHFTTIVYQTTMSSGGAIMGEDPKTSVAVNRYLQ 563
 QY 487 VMDFANLYVAGNGTIRGFGENPT 510
 DB 564 SWDVNVFVPGASAPFQGLGYNPT 587

RESULT 9

Q9HQR8 PRELIMINARY; PRT; 529 AA.
 AC Q9HQR8;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Vng1035C.
 GN VNG1035C.
 OS Halobacterium sp. (strain NRC-1).
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
 OC Halobacteriaceae; Halobacterium.
 OX NCBI_TaxID=64091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20504483; PubMed=11016950;
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
 RA Shukla H.D., Lasky S.R., Balliga N.S., Thoreson V., Sbrogna J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
 RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,

OS TaxID=305;
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea;
OX NCBI_TaxID=53336;
PN (1)

RP SEQUENCE FROM N.A.
 RC STRAIN=1056R;
 RX MEDLINE=20200361; PubMed=10735866;
 RA Pujol C.J., Kado C.I.;
 RT "Genetic and Biochemical characterization of the pathway in Pantoea
 citrea leading to pink disease of pineapple.";
 RL J. Bacteriol. 182:2230-2237(2000).
 DR EMBL; AF131202; AAD44706.1; -
 DR InterPro; IPR000205; NAD_binding.
 SQ SEQUENCE 553 AA; 59939 MW; F2A60424416B8C3F CRC64;

Query Match 5.1%; Score 148; DB 2; Length 553;
 Best Local Similarity 20.8%; Pred. No. 0.0012;
 Matches 114; Conservative 58; Mismatches 222; Indels 160; Gaps 21;

QY 52 LDGCAWAPGSSAIGNKNPHOREFENLSAEAVTRGVGMSHTWCSTPRHPP---NE 108
 DB 81 LAPAPLYPPNNYV---NVTGPSAGSQ---QGYLTVGGTTWHWAASCRHHPSPFVVK 134
 QY 109 SLPGIGRP-KLSNDPAEDDKWENLYSEARLIGT-----STKEFDESIRHTLVLR 158
 DB 135 SKYGVGRDWPISYDEME---PW---YCEAEYIGVAGSPDSMQSPSERSRPYPMVMPF 188
 QY 159 SLQDAY-----KDRQIRFP---LPLAC-----HRLKN 183
 DB 189 AHGDTYFASVYVPHGYNLVPIPOGRSTRPWGRPVCCGNNNCOPICPGAMYNIGHIER 248
 QY 184 APEYVEMHSAENLFHSIYNDKOKLFTLLTNHRCR---LALTGGYKKGIGAAEVRNLL 240
 DB 249 AESKGVAVLAESVYKIDTDD-----NNRVTAVHWLDNQGASHKATG----- 290
 QY 241 ATRNPPSQLDSYIMAKVYVLAAGAIGNPOILYNSGFGSLQVTPRNDLSLPN-----LGR-- 294
 DB 291 -----KAFALACNGIETPRLLQA-----ANKANPTGIANSDDMVGRNN 329
 QY 295 -----YITQPMAFQIVLRQEFVDSVRDPPGLPWKEAVAQHIATKPTDALPIP 345
 DB 330 MDHSGFHCSELTPEPVNLGRGPAQSSCMVGRDGAFRSEY---SANKMILNISRVPAT 386
 QY 346 FRDPEQVTPFTTEHPHWTQIHRDAFSYGAVPEVDSRVI-----VDLRWFGATDPEA 399
 DB 387 -----KQALAKGLVKALDEIRYRSIHGVLSLSLEPLPD 423
 QY 400 NNLLVFQNDVQDYSMPQPTFRYPSTASNVARKMADMCEVASNLGGYLPSTPPQFMD 459
 DB 424 ENRLTLKTRKDPHGLACPDIDYVDGYVRKGATAAHEQLQHLGSLFNG-----KEFN 476
 QY 460 PGLAL-----HLAGTTRIGFKATTVADNLSLVDFANLYVAGNGTIRTFGFGNPTILSMC 515
 DB 477 ITTALNANNHMGTTGSKADAVVDGNCRTFDHENLWLPUGGAIPSASVYVNSTLSMAA 536
 QY 516 HAKSARSILNLIK 529
 DB 537 LGLKAHDIISLRMK 550

RESULT 14

Q9R226 PRELIMINARY; PRT; 722 AA.
 ID Q9R226
 AC Q9R226;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 OC Deinococcus-Thermus; Deinococci; Deinococcales;
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE GMC oxidoreductase.
 GN DRA0127.
 OS Deinococcus radiodurans.
 OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
 OC Deinococcaceae; Deinococcus.
 OC NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R1;
 RX MEDLINE=20036896; PubMed=10567266;

RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Croesby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Uterback T., Zalewski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium Deinococcus
 radiodurans R1.";
 RL Science 286:1571-1577(1999).
 DR EMBL; AE001862; AAF12230.1; -
 DR TIGR; DRA0127; -
 DR InterPro; IPR001064; Crystallin.
 DR InterPro; IPR001199; Cyt_B5.
 DR InterPro; IPR000172; GMC_oxred.
 DR Pfam; PF00732; GMC_oxred; 1.
 DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
 DR PROSITE; PS00191; CYTOCHROME_B5_1; 1.
 DR PROSITE; PS00624; GMC_OXRED_2; 1.
 DR Complete proteome.
 SQ SEQUENCE 722 AA; 75376 MW; 450DFICD1B7596F9 CRC64;

Query Match 5.0%; Score 144; DB 16; Length 722;
 Best Local Similarity 21.1%; Pred. No. 0.004;
 Matches 107; Conservative 68; Mismatches 196; Indels 136; Gaps 26;

QY 90 GGMSTHTCSTPRIHPMESLPICRKLNS-----DPAEDDKWENLYSEARLIG 141
 DB 277 GGSVTWNSV---PPRDDI---QRWASEHGLSDVADPGVD---RHIDAVLRM-G 323
 QY 142 TSTKEFDESIRHTLVRLSDAYKQRIFRPLPLACHLKNAPVEVHSAENLFHSIY 201
 DB 324 VSEQCSHDNGPHQ---RLVEGADKLGYTEFKA-----ALNLSPEH---YDADKAGHAGF 371
 QY 202 NDD---KOKKLP-----LLTNHRCRLALTGGYKKGIGAAEVRNLLATRNPS 247
 DB 372 GDQTAQKQTLNFTLKDAFEAGARILVGTARQVRLVEDG-----RAAGVSAVTMGDETR 426
 QY 248 QLDYSIMAKVYVLAAGAIGNPOILYNSGFGSLQVTPRNDLSLPNGLRYITEQPMAFQIV 307
 DB 427 QIT--VRAPQVVVACALETPALLRSIGG-----PAAGRYLRLHPAGLVAGI 473
 QY 308 LRQEFVDSVRDDPYG---LPWKEAVAQHIANKPTDALIPRDPE-----P 351
 DB 474 -----YGEDQRAMWP-----POSGILKQFADHENGHGFIEGVQYGP 511
 QY 352 QVTT---PPT---EEH-PWHTQIHRDAFSYGAVPEVDSRVIVD-----LRWFGATDP-EA 399
 DB 512 ALMASGLPMTGGAHRDLMSKFRHATFVSIYQDRGHGQVTVVDEGNVHTYALTDDDA 571
 QY 400 NNLLVFQNDVQDYSM-----PQPTFRYPSTASNVARKMADMCEVAS---NLGGYLP 451
 DB 572 RN---FRRGVTEISIRLHEAAGAEIIVALAPGVPWRERGGDLBAFIGQAVPLGAGQTV 628
 QY 452 TSPQFMDPGLALHLAGTTRIGFKATTVADNLSLVDFANLYVAGNGTIRTFGFGNPTIL 511
 DB 629 FS-----AHQWGSARMGSDPQTSVADPDQLHVDVPGVWIGTSAFPTCSGVNPMV 678
 QY 512 TSMCHAIKARSSTLTKGGTDGKNTG 538
 DB 679 SCMALASRTAEKLLAMWEGADGTGSG 705

RESULT 15

Q98C76 PRELIMINARY; PRT; 523 AA.
 ID Q98C76
 AC Q98C76;
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Hypothetical protein mlr5266.
 GN MLR5266.
 OS Rhizobium loti (Mesorhizobium loti).

```

search completed: October 28, 2003, 15:59:49
job time : 109 secs

```